Thermionic Electron Emission *

By J. A. BECKER

Introduction

THERE have appeared in the *Reviews of Modern Physics* two excellent summaries on thermionic emission, one by Compton and Langmuir [1] and one by Dushman.[2] Compton and Langmuir, while dealing primarily with discharge in gases, also discussed many phases of thermionic emission. Dushman's article is a comprehensive review on thermionics. He faithfully reflects whatever viewpoints and experiments appear in the literature. Besides reviewing the work that has been performed since 1930, the present article will be an attempt to review in a critical manner some of the matters which in the preceding reviews were left undecided. On the other hand, no attempt will be made to give a complete presentation of all the views appearing in the literature. As to the close connection between thermionic and adsorption phenomena, this will be dealt with in an article now in preparation.

Recently there have been published two comprehensive books on thermionics. One is in English by A. L. Reimann.[3] The other is Vol. IV of Müller-Pouillets *Lehrbuch der Physik* [4] edited by A. Eucken, with contributions by A. Eucken, R. Suhrmann, L. Nordheim and others. The topics which are fully covered in these two books and in the book [5] by W. Schottky and H. Rothe, *Physik der Glühelektroden* will not be covered in detail in the present article. Since photoelectric phenomena are closely associated with thermionics, it is well to refer also to Linford's [6] review on the external photoelectric effect and the book by Hughes and DuBridge [7] on photoelectric phenomena.

Empirical and Theoretical Richardson Formulæ

One topic on which considerable confusion has existed goes to the very root of thermionic emission. It is the interpretation that is to be put on the slope and intercept of a Richardson line and how the slope and intercept are related to certain quantities in theoretical formulæ. Empirically it is found that the thermionic emission current density, i, is related to the temperature, T, by the Richardson formula

$$i = A_n T^n \exp(-b_n/T), \tag{1a}$$

* Published in *Reviews of Modern Physics*, April, 1935.

or its equivalent

$$\log_{10} i = \log_{10} A_n + n \log_{10} T - (b_n/2.3T). \tag{1b}$$

A_n and b_n are constants characteristic of the surface. Their value depends on the value assigned to n. From such experiments it is impossible to decide whether n should equal 0 or 4 or any value between these. There are good theoretical reasons, which are given below, why $n = 2$. In that case

$$i = AT^2 \exp(-b/T), \tag{2a}$$

or

$$\log i - 2 \log T = \log A - b/2.3T. \tag{2b}$$

If $\log i - 2 \log T$ is plotted *versus* $1/T$, a straight line is usually obtained. Call this line a Richardson line. Its slope is $-b/2.3$, and its y intercept is $\log A$. Since we shall have numerous occasions to refer to the slope and intercept of a Richardson line, we shall find it convenient to refer to them by their equivalents $-b/2.3$ and $\log A$, respectively. On those rare occasions when the Richardson plot yields a curved line, we can draw a tangent at any point on the curve. Equation (2) will then represent the equation for this tangent; $-b/2.3$ and $\log A$ will depend on the particular point at which the tangent is drawn, so that b and A will depend on T.

The Thermodynamic Equation

The slope and intercept of experimental Richardson plots are to be correlated with certain quantities in one or the other of two theoretical equations. The first of these * is based on the first and second law of thermodynamics and the assumption that the electron vapor acts like a perfect gas.† The equation is:

$$\log i_T = \log i_{T'} + \log [(1 - r)/(1 - r')] + \tfrac{1}{2} \log T' - \tfrac{1}{2} \log T + (1/2.3) \int_{T'}^{T} (L_p/RT^2) dT, \tag{3}$$

in which T' is any fixed temperature in the experimental temperature range; r and r' are the electron reflection coefficients at T and T', respectively; L_p is the heat of vaporization per g. mole of electrons at constant pressure; R is the gas constant per g. mole.

Thermodynamics cannot tell us how L_p varies with T and until we know this we cannot perform the integration indicated. By consider-

* For a recent critical derivation see Becker and Brattain.[8]

† This assumption is subsequently justified by experiment.

ing the mechanism by which the electrons evaporate from the metal, we can arrive at some conclusions regarding the temperature dependence of L_p. Since in the derivation of equation (3) it was assumed that the electron vapor acts like a perfect gas, it follows that when 1 g. mole of electrons is vaporized at constant pressure an amount of work RT must be done against the external pressure and an amount of heat $(3/2)RT$ must be provided to furnish the known mean kinetic energy of the vaporized electrons. It then becomes desirable to define a new quantity h by the equation,

$$h = (L_p/R) - (5/2)T. \tag{4}$$

Since h plays an important role in the final formula, it will be convenient to give it the name "heat function." * The product kh, where k is Boltzmann's constant, represents the average heat of vaporization per electron less $(5/2)kT$. Substituting equation (4) in equation (3),

$$\log i = \log i_{T'} + \log [(1 - r)/(1 - r')] - 2 \log T' + 2 \log T + (1/2.3)\int_{T'}^{T} (h/T^2)dT. \tag{5}$$

Thermodynamics alone cannot tell us how the heat function h varies with T and we cannot perform the indicated integration until this is known. However, we can deduce an important theorem even without performing the integration: *If the experimental value of log i − 2 log T is plotted versus $1/T$, the slope of the tangent at any value of T is* $-h/2.3$.†

Hence for those surfaces for which the Richardson lines are straight, h is independent of T in the experimental range. For these surfaces, equation (5) reduces to

$$\log i = \log i_{T'}/(1 - r')(T')^2 + h/2.3T' + \log (1 - r) + 2 \log T - h/2.3T = \log H(1 - r) + 2 \log T - h/2.3T, \tag{6}$$

where

$$\log H = \log i_{T'}/(1 - r')(T')^2 + h/2.3T'.$$

Log $H(1 - r)$ is the intercept of the Richardson line on the y axis.

An alternative derivation of the thermodynamic emission equation uses the absolute zero of temperature as the lower limit in the various integrals. In this way Bridgman derives the equation,‡

$$i = U\alpha(1 - r)T^2 \exp. [- L_0/kT + \varphi(T)], \tag{6a}$$

* This is of course not the heat function used in thermodynamics. The heat function defined here has the dimensions of temperature. It is often given in volts $V = kh/e$. Later h will also be used for Planck's constant but we believe no confusion will arise.

† For the proof see Becker and Brattain.[8] In the proof it is assumed that dr/dT is zero or very small. This assumption is justifiable.

‡ See Eq. IV, 33 on page 99 of his book named in References.[9]

in which U is a universal constant equal to $2\pi Gk^2me/h^3 = 120$ amps./cm.^{2} °K.^{2}; h (Planck's constant), m, e and k have the customary significance; G is the statistical weight which is equal to 2 for electrons; $\alpha = \exp(S_{0\rho}/k)$; $S_{0\rho}$ is the entropy per atom of a metal whose surface has a charge density ρ at $T = 0$; L_0 is the heat of vaporization per electron at constant pressure at $T = 0$;

$$\varphi(T) = \frac{1}{k}\int_0^T \frac{(C_{p\rho} - C_{pm})}{T}\,dT - \frac{1}{kT}\int_0^T (C_{p\rho} - C_{pm})dT;$$

$C_{p\rho}$ is the specific heat per atom at constant pressure when the metal surface has a charge density ρ while C_{pm} is the specific heat for the uncharged metal. In the derivation it is assumed that the entropy of the uncharged metal at $T = 0$ is zero in accordance with the third law of thermodynamics; it is also assumed that the electron vapor acts like a perfect gas. The value of U follows from the value of the entropy constant of a perfect gas deduced from quantum statistics. Up to the present time neither theory nor experiment has yielded numerical values for α or $\varphi(T)$. If, however, it is assumed that $\alpha = 1$, $\varphi(T) = 0$ and $r = 0$ then equation (6a) reduces to

$$i = UT^2 \exp.(-L_0/kT), \tag{6b}$$

which is the equation derived by Dushman [10] in 1923. It predicts that all Richardson lines should have the same intercept on the y axis, namely, log U. Since this prediction is not fulfilled by experiment it would appear that the assumptions made in obtaining equation (6b) are not valid. It may be well to point out also that adsorbed particles on the surface probably contribute additional terms to the expression for the specific heats and entropy at absolute zero. These have not been taken into account.

We are now in a position to show why the exponent of T in equation (1a) should be 2. To do this we consider h in equation (5) or L_p in equation (3). The heat of vaporization L_p is defined as the heat energy that must be added to the system in order to evaporate one g. mole of electrons at constant pressure. We have seen that $(5/2)RT$ ergs must be added to account for the specific heat of the vaporized electrons and work done against the external pressure. The remainder, Rh, which includes all other energies can be put equal to $\bar{P} - \bar{K} - T(d\bar{P}/dT)$ where $\bar{P}$ is the increase in potential energy of the electrons, and $\bar{K}$ is the mean kinetic energy which the electrons had in the metal. $\bar{P}$ includes work done against the image force or any other electrical forces. So little is known about the exact nature of $\bar{P}$

that there is little point in examining the temperature dependence of the quantity $\bar{P} - Td\bar{P}/dT$ more closely. On the other hand, the quantity $\bar{K}$ and its variation with temperature does depend on the particular assumption that is made with regard to the energy distribution of the free electrons in the metal. In particular $\bar{K}$ is very nearly independent of T if the electrons in the metal have kinetic energies given by the Fermi-Dirac function.* That this is the correct distribution function is quite well established by the numerous successes which this theory has had in explaining experimental facts in connection with metals.[11–16] For the Fermi-Dirac distribution $\bar{K}$ is practically a constant term in the expression for the heat function h. There is then no reason for changing the form of equation (6) which contains the term 2 log T. This is equivalent to an exponent of 2 in equation (1a).

The case was somewhat different before the advent of the quantum theory. The electrons in the metal were then assumed to act like a perfect gas. Hence the energy $\bar{K}$ was taken to be $(3/2)RT$. It was thus natural to subtract this from the $(5/2)RT$ for the electron vapor. In this way one is led to an expression for log i similar to equation (6), but instead of 2 log T there now appears $\frac{1}{2}$ log T. So that the exponent of T in equation (1a) was taken to be $\frac{1}{2}$.

It is well to note that on the basis of this thermodynamic argument, there is no good reason why the heat function should be independent of T and why the Richardson lines should be straight. Experiment shows, however, that for nearly all surfaces which are not close to their melting point, the heat function is independent of T to within experimental error. In the neighborhood of the melting point, the heat function varies with T. It should also be noted that thermodynamics does not predict that all Richardson lines should have a common intercept on the y axis. This prediction which is true only for special classes of surfaces has been made on the basis of a statistical theory which we will now discuss.

The Statistical Equations

a. Classical treatment. If we knew the velocity distribution and density of the electrons inside a metal at various temperatures and the difference in potential energy between an electron at rest inside and outside the metal, it would be a comparatively easy task to determine statistically how many electrons could escape from a square centimeter of surface in one second. It was at first assumed that the electrons inside the metal acted like a perfect gas; the velocity distribution is

* This function will be discussed later in this paper.

given by Maxwell's law

$$n(u, v, w)dudvdw = \frac{nm^{\frac{3}{2}}}{(2\pi kT)^{\frac{3}{2}}} \exp. \left[\frac{-m(u^2 + v^2 + w^2)}{2kT} \right] dudvdw, \quad (7)$$

where u, v, w are the velocities in the x, y, z directions respectively; $n(u, v, w)$ is the number of electrons per cm.^{3} having u values in the range (u, du), i.e., between u and $u + du$, v values in the range (v, dv), and w values in the range (w, dw); n is the total number of electrons per cm.^{3}; m is the electron mass; and k is Boltzmann's constant. The number of electrons having u components of velocities in the range (u, du) is obtained by integrating equation (7) with respect to v and w from $-\infty$ to $+\infty$.

$$n(u)du = n(m/2\pi kT)^{\frac{1}{2}} \exp. (-mu^2/2kT)du. \quad (8)$$

The total number of particles striking a cm.^{2} of surface per second is given by

$$N_t = \int_0^\infty un(u)du = n(kT/2\pi m)^{\frac{1}{2}}. \quad (9)$$

But only those electrons whose values of u exceeds $u_0 = (2p/m)^{\frac{1}{2}}$ will escape from the surface. The quantity p, called the work function, represents the potential energy of the electron outside the metal; it is the work that must be done to take an electron at rest in the metal and transport it across the surface to a distance at which the surface forces are negligible. The total number, N, of electrons which can escape from one cm.^{2} of surface in one second is then

$$\begin{aligned} N &= \int_{u_0}^\infty un(u)du \\ &= \int_{u_0}^\infty n(m/2\pi kT)^{\frac{1}{2}} u \exp. (-mu^2/2kT)du \\ &= n(kT/2\pi m)^{\frac{1}{2}} \exp. (-p/kT). \end{aligned} \quad (10)$$

The emission current in amperes per cm.^{2} is

$$i = Ne = A'T^{\frac{1}{2}} \exp. (-p/kT) \quad (11a)$$

or

$$\log i - \tfrac{1}{2} \log T = \log A' - p/2.3kT, \quad (11b)$$

where

$$A' = ne(k/2\pi m)^{\frac{1}{2}}, \quad (12)$$

e = charge on the electrons in coulombs = 1.59×10^{-19}.

If p is independent of T and if $\log i - \frac{1}{2}\log T$ is plotted *versus* $1/T$, the slope $= -p/2.3k$ and the y intercept $= \log A'$ or $\log ne(k/2\pi m)^{\frac{1}{2}}$. For clean tungsten it is found by experiment that the current density can be represented by

$$i = 2.06 \times 10^7 T^{\frac{1}{2}} \exp.\ (-55{,}300/T).$$

From this it follows that $ne(k/2\pi m)^{\frac{1}{2}} = 2.06 \times 10^7$ and that $n = 8.4 \times 10^{20}$ electrons/cm.^{3} This is to be compared with 635×10^{20} atoms/cm.^{3} So that if we postulate one "free electron" for every 75 atoms, we can account for the observed thermionic emission classically.

Such a concentration of free electrons may be considered to be in quite good accord with the first of two possible deductions from experiments on specific heats. From these it follows that: (1) Either the number of free electrons must be small compared to the number of atoms and the mean kinetic energy per electron is $(3/2)kT$; or else, (2) the number of free electrons is of the order of the number of atoms but the kinetic energy increase per degree rise in temperature is much smaller for electrons than it is for atoms. The correlation of experiment and classical theory in the case of the optical properties, electrical conductivity, thermoelectricity, Thomson and Peltier effects lead to certain inconsistencies. These disappear when theories based on the Fermi-Dirac distribution are used for these effects and it is postulated that the number of free electrons in metals is of the same order as the number of atoms. The classical theory for Richardson's equation thus leads to values of n which are incompatible with values deduced from these effects. The newer theory has also made progress in explaining ferromagnetism. It is thus a better basis for a statistical theory of electron emission. Such a theory was developed by Sommerfeld [11] and Nordheim.[16]

b. Quantum-mechanical treatment. The Fermi-Dirac theory gives the velocity distribution as

$$u(u, v, w)dudvdw = \frac{Gm^3}{h^3} \times \frac{1}{M^{-1}\exp.\,[m(u^2 + v^2 + w^2)/2kT] + 1} dudvdw, \quad (13)$$

G is the statistical weight; for electrons its value is 2. h is Planck's constant. The quantity M is so adjusted that the integral of $n(u, v, w)$ gives the total number of electrons/cm.^{3} This integration is so difficult that no relatively simple and exact expression for M can be found.

However, in two limiting cases good approximations have been obtained.

In the first case M is so small a quantity that

$$M^{-1} \exp. [m(u^2 + v^2 + w^2)/2kT] \gg 1.$$

It then follows that

$$M = nh^3m^{\frac{3}{2}}/G(2\pi kT)^{\frac{3}{2}} \tag{14}$$

and that equation (13) is the same as equation (7) for the classical treatment.

In the second case M is a large quantity and the 1 in the denominator of equation (13) cannot be neglected. Sommerfeld [11] has shown that in this case

$$M = \exp. (K/kT), \tag{15}$$

where

$$K = \frac{h^2(3n/4\pi G)^{\frac{2}{3}}}{2m}\left[1 - \frac{(2\pi mkT)^2}{12h^4}\left(\frac{3n}{4\pi G}\right)^{-4/3} + \cdots\right]. \tag{16}$$

The second term in the brackets is usually a very small numerical quantity and can nearly always be neglected.

If we assume that n, the number of electrons/cm.^{3}, is equal to the number of atoms/cm.^{3} in a metal or a small factor times this number, we can compute M for case 1 by equation (14) or for case 2 by equation (15). In either case M turns out to be a large quantity. Hence for metals the second case is applicable while the first case is not. Hence

$$n(u, v, w)dudvdw = \frac{Gm^3}{h^3} \times \frac{1}{\exp.\left[\frac{\frac{1}{2}m(u^2 + v^2 + w^2) - K}{kT}\right] + 1} dudvdw, \tag{17}$$

since

$$M^{-1} = \exp. (-K/kT).$$

Integrating this from $-\infty$ to $+\infty$ with respect to v and w Nordheim [14] has shown that the number of electrons per cm.^{3} having velocities in the range (u, du), i.e., between u and $u + du$, is

$$n(u)du = \frac{2\pi Gm^2kT}{h^3} \ln\left[1 + \exp.\left(\frac{K - \frac{1}{2}mu^2}{kT}\right)\right] du. \tag{18}$$

The number of electrons striking a surface normal to the u direction per cm.^{2} per sec. and having velocities in the range (u, du) is given by

$N(u)du = un(u)du$; hence

$$N(u)du = \frac{2\pi Gm^2kT}{h^3} u \ln\left[1 + \exp.\left(\frac{K - \frac{1}{2}mu^2}{kT}\right)\right] du. \quad (19)$$

Now only those having velocities greater than u_c will be able to cross the surface and escape where u_c is given by $\frac{1}{2}mu_c^2 = P_m$. P_m is the difference in potential energy between an electron at rest inside and outside the metal. Now P_m is about 3/2 times as large as K and therefore $\frac{1}{2}mu_c^2 > 1.5K$. Also for the values of T encountered in thermionic experiments kT is small compared with $(\frac{1}{2}mu_c^2 - K)$. Therefore, for values of $u > u_c$, exp. $[(K - \frac{1}{2}mu^2)/kT]$ is a very small quantity and

$$\ln\left[1 + \exp.\left((K - \tfrac{1}{2}mu^2)/kT\right)\right] = \exp.\left[(K - \tfrac{1}{2}mu^2)/kT\right]$$

to a good approximation. This follows, since

$$\ln(1 + \Delta) = (\Delta - \tfrac{1}{2}\Delta^2 + \tfrac{1}{3}\Delta^3 - \tfrac{1}{4}\Delta^4 + \cdots)$$

provided $\Delta^2 < 1$. Hence, for $u > u_c$ and the temperatures encountered in thermionic emission

$$N(u)du = \frac{2\pi Gm^2kT}{h^3} u \exp.\left(\frac{K - \frac{1}{2}mu^2}{kT}\right) du. \quad (20)$$

The number that cross the surface per cm.^{2} per second is given by

$$N = \int_{u_c}^{\infty} N(u)du = \frac{2\pi Gm^2kT}{h^3} \int_{u_c}^{\infty} u \exp.\left(\frac{K - \frac{1}{2}mu^2}{kT}\right) du$$
$$= \frac{2\pi Gmk^2T^2}{h^3} \exp.\left(\frac{K - \frac{1}{2}mu_c^2}{kT}\right)$$
$$= \frac{2\pi Gmk^2}{h^3} T^2 \exp.\left(-\frac{P_m - K}{kT}\right). \quad (21)$$

Finally

$$i = Ne = (2\pi Gmek^2/h^3)T^2 \exp.\left[-(P_m - K)/kT\right]$$
$$= UT^2 \exp.(-W/kT) = UT^2 \exp.(-w/T)$$
$$= UT^2 \exp.(-\varphi e/kT). \quad (22)$$

where

$$U = 2\pi Gmek^2/h^3 \quad (23)$$

and

$$P_m - K = W = kw = \varphi e. \quad (24)$$

If i is expressed in amperes per cm.^{2}, the value of U is 120 amperes/cm.^{2} °K.^{2} The quantities W, w, or φ are called the work function; the difference between them is merely one of units. φ is expressed in volts, w in degrees Kelvin, and P_m, K and W in ergs. P_m is called the outer work function and K, the inner work function. Oftentimes it is convenient to refer to P_m, K and W as if they were expressed in volts.

c. Treatment in terms of energies. For many purposes it is convenient to have expressions for the distribution in energies instead of in velocities. We can then express these energies in equivalent volts and obtain numerical values which are more familiar. Let

$$E = (m/2)(u^2 + v^2 + w^2); \qquad E_n = (m/2)u^2,$$

" the normal component of the energy "; $V_n = E_n/e$; $n(E)dE =$ the number of electrons per cm.^{3} having energies in the range (E, dE); similarly for $n(E_n)dE_n$; $N(V_n)dV_n$ is the number striking a cm.^{2} of surface per second having normal component of energies in the range (V_n, dV_n).

Then

$$n(E)dE = \frac{2\pi G}{h^3}(2m)^{\frac{3}{2}} \frac{E^{\frac{1}{2}}}{1 + \exp.\,[(E - K)/kT]} dE, \tag{25}$$

$$n(E_n)dE_n = \frac{2^{\frac{1}{2}}\pi G m^{\frac{1}{2}} kT}{h^3} \frac{1}{E_n^{\frac{1}{2}}} \ln\left[1 + \exp.\frac{(K - E_n)}{(kT)}\right] dE_n, \tag{26}$$

$$N(E_n)dE_n = \frac{2\pi G m k T}{h^3} \ln\left[1 + \exp.\left(\frac{K - E_n}{kT}\right)\right] dE_n, \tag{27}$$

$$N(V_n)dV_n = \frac{2\pi G e m k T}{h^3} \ln\left[1 + \exp.\left(\frac{K - V_n e}{kT}\right)\right] dV_n. \tag{28}$$

Equations (25) and (26) are readily derived from equation (18); while equations (27) and (28) follow from equation (19). It is also instructive to compare equation (28) with the corresponding equation which is based on classical statistics, namely,

$$N(V_n)dV_n = n(e^2/2\pi mkT)^{\frac{1}{2}} \exp.\,(-V_n e/kT)dV_n. \tag{29}$$

This is readily derived from equation (8).

d. Comparison between classical and quantum-mechanical treatment. Comparison between equations (28) and (29) is best brought out by a graph such as Fig. 1 which shows log $N(V_n)$ *versus* V_n for the two cases. It is to be remembered that $N(V_n)dV_n$ is the number of electrons in the metal which strike 1 cm.^{2} of surface per second whose

energy components normal to the surface are in the range (V_n, dV_n) volts. It has been customary to plot $N(V_n)$ *versus* V_n for equation (28). At $T = 0$, $N(V_n)$ decreases linearly with V_n from a value of

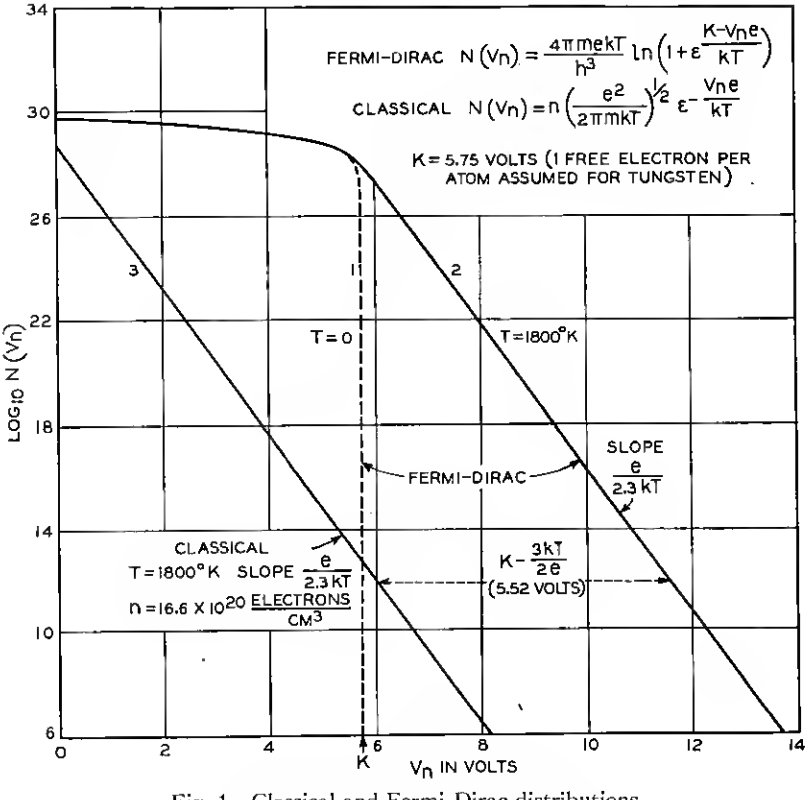


Fig. 1—Classical and Fermi-Dirac distributions.

$2\pi Gemk/h^3$ when $V_n = 0$, to zero when $V_n = K/e$; for $V_n > K/e$, $N(V_n) = 0$. For $T > 0$, the function is much the same except in the neighborhood of $V_n = K/e$ and for $V_n > K/e$; the curve is here everywhere higher than the curve for $T = 0$ and decreases exponentially. Since only those electrons can escape for which $V_n \geq P_m > (3/2)K$, we are primarily interested in the exponential portion of the curve. It is therefore more advantageous to plot $\log N(V_n)$ rather than $N(V_n)$.

In Fig. 1 curves 1 and 2 are for equation (28) at $T = 0$ and $T = 1800°$ K., respectively; while curve 3 is for the classical case or equation (29). For curves 1 and 2, the value of K/e has been taken as 5.75 volts which is the value appropriate for tungsten assuming one

free electron per atom. For curve 3 the value of n has been so chosen that this curve is shifted with respect to curve 2 by $K/e - 3kT/2e$ or 5.52 volts. The value of n which does this is 16.6×10^{20}/cm.^{3} To account for the observed emission from tungsten we have previously deduced a value $\frac{1}{2}$ as great or 8.4×10^{20}/cm.^{3} The factor of 2 is due to the fact that the intercept of the observed Richardson plot for tungsten corresponds to 60 amp./cm.^{2} °K^2 while the theoretical intercept corresponds to 120 amp./cm.^{2} °K.^{2}

At first sight it might appear that the shift between curves 2 and 3 should be K/e rather than $K/e - 3kT/2e$. The additional term is accounted for by comparing the classical or $T^{\frac{1}{2}}$ equation (11a) with the quantum-mechanical or T^2 equation (22). It is well known that the experimental results can be made to fit either the $T^{\frac{1}{2}}$ or the T^2 equation and that the constants in the two equations are related by

$$W(\text{or } P_m - K) = p - (3/2)kT, \tag{30}$$

and

$$A = A'/\epsilon^{\frac{1}{2}}T^{\frac{1}{2}}. \tag{31}$$

From equation (30) it follows that the classical work function p is larger than the quantum-mechanical work function W or $P_m - K$ by $(3/2)kT$ and that to obtain the same emission from the two distributions the curves must be shifted by $K/e - 3kT/2e$.

The Temperature Dependence of the Work Function

Thus far little has been said about the temperature dependence of the work function. While there is no good theoretical reason for expecting a large temperature dependence, there is also no good reason to expect that the work function is accurately independent of T. Experiments on contact potential and photoelectric effect indicate that there is indeed a small temperature effect.* In investigating the effect of the temperature dependence we shall limit ourselves to the quantum-mechanical equations. However, a similar treatment would be applicable to the classical or $T^{\frac{1}{2}}$ equation.

If in equation (22), w or its equivalents W or φ are independent of T, then the slope of a Richardson line is $-w/2.3$ or $-W/2.3k$ or $-\varphi e/2.3k$; the intercept is log U. So that

$$b = w = W/k = \varphi e/k \quad \text{and} \quad \log A = \log U. \tag{32}$$

If the work function varies linearly with temperature,

$$w = w_0 + \alpha T \quad \text{or} \quad W = W_0 + \alpha kT$$

* For a detailed discussion see reference 8.

or

$$\varphi = \varphi_0 + \alpha(k/e)T; \tag{33}$$

where $\alpha = dw/dt$ is a constant independent of T; its units are degrees per degree. The slope of a Richardson line is now $-w_0/2.3$ so that

$$b = w_0 = W_0/k = \varphi_0 e/k$$

while

$$\log A = \log U - \alpha/2.3. \tag{34}$$

Since the slope is constant, the Richardson line is straight. This line is determined either by the empirical constants A and b or by the values of w and dw/dT in theoretical equations.

If w is a general function of T, the Richardson line will be curved. If a tangent is drawn at a point corresponding to any temperature, the slope of the tangent is $-(1/2.3)(w - Tdw/dT)$ and its intercept is $\log U - (1/2.3)dw/dT$; w and dw/dT are to be taken at the point of tangency. Hence

$$b = w - Tdw/dT$$

and

$$\log A = \log U - (1/2.3)dw/dT. \tag{35}$$

In a previous section it was shown that the slope of the Richardson line is always equal to $-h/2.3$. Hence

$$-h/2.3 = -(1/2.3)(w - Tdw/dT)$$

or

$$h = w - T(dw/dT). \tag{36}$$

This important equation gives the relation between the heat function and the work function. It is similar in form to the relation between the total energy E and the free energy F, *viz.*,

$$E = F - T(dF/dT). \tag{37}$$

The distinction between the heat function h and the work function w is strikingly brought out in Fig. 2. The slope of the Richardson line is $-h/2.3$, while the slope of a straight line connecting any point on the Richardson line with the intercept $\log U(1 - r)$ is $-w/2.3$.

The theory that the work function is indeed a function of temperature has been championed in recent times by R. Suhrmann and his collaborators. A good account of this work can be found in Volume 4 of Müller-Pouillets *Lehrbuch der Physik*. One method by which Suhrmann has shown the temperature dependence of the work function is

that of the complete photoelectric emission. The surface to be investigated is illuminated by light from a source whose temperature is varied. It is found that the resulting photo-current obeys a Richardson law and the slope of the Richardson line is taken as the work

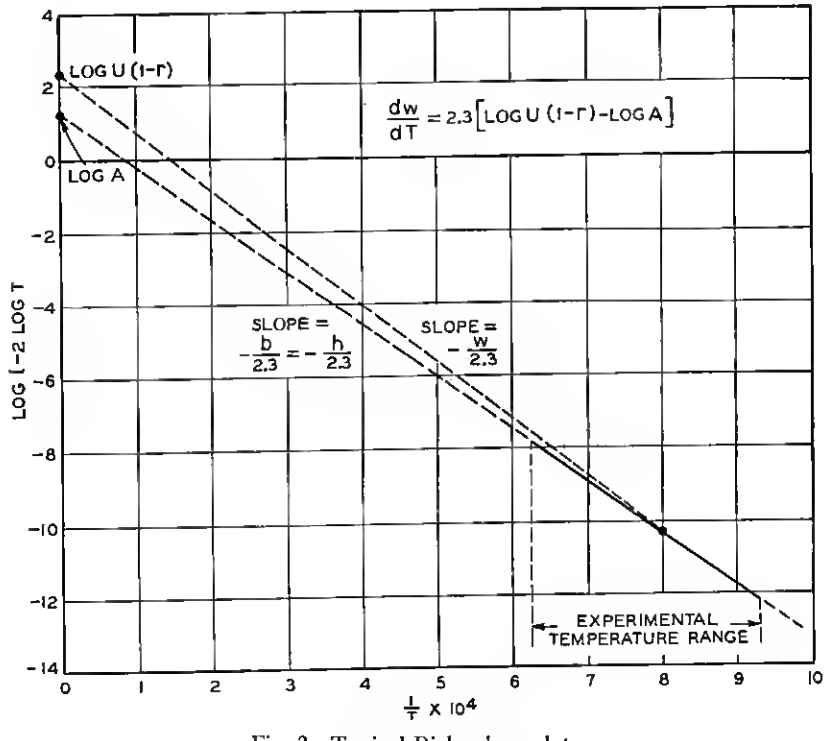


Fig. 2—Typical Richardson plot.

function. The temperature of the cathode is now altered and a new Richardson line is obtained. It is found that the slope has changed. It may be worth while to analyze critically the theory of this experiment to ascertain whether the slope is related to the work function or to the heat function.

On True and Apparent Surface Areas

One other point in the correlation between experiment and theory is to be noted. In the empirical equation, i is the current per cm.^{2} of apparent surface; while in the theoretical equation, i is the current per cm.^{2} of an ideal or true surface. The real surface in thermionic experiments is not smooth; it usually consists of a large number of

etch facets which are oriented at various angles with respect to a mean surface plane. The appearance of the surface is to be compared with an airplane view of a city whose gabled roofs have various designs and various angles. The size and shape of the etch facets depend on the material of the cathode, the crystal size, the orientation of the crystal with respect to the mean surface, the degree of heat treatment and presumably some unknown factors. Theoretically it is possible to deduce values of S, the ratio of the true surface area to the apparent surface area, for certain simple cases. Thus, Tonks [17] has computed the following average values of S: For cubic facets or 100 planes, 1.500; for dodecahedral facets or 110 planes, 1.225; for 100 and 110 planes, 1.129. Some of the assumptions on which these values are based are: (1) The surface is covered with pyramids whose sides are crystal planes; (2) the orientation of crystal axes with respect to the surface is random; (3) for a given type of etch plane or planes, the facets occur in such a way as to give a minimum surface area. No one has made a thorough investigation to test these assumptions by experiment. Some microscopic pictures of etched surfaces which I have seen showed truncated pyramids in contrast with the first assumption; they also showed sub-facets, thus violating the third assumption.

Values of S have been obtained from experiments on adsorption of gases on solid and liquid surfaces. Particularly significant experiments are those of Bowden and Rideal [18] on the adsorption of hydrogen ions deposited on metal surfaces by electrolysis of a solution of sulphuric acid. The potential of these surfaces was determined against a calomel electrode. They found that when the electrolytic current exceeded a minimum value, the surface potential increased linearly with the quantity of electricity until it reached a new steady value. For a mercury surface as well as for a thin film of platinum on mercury the potential increased by one volt for 6×10^{-6} coulomb/cm.^{2} The direction of the potential change and its amount are such as to be expected if hydrogen ions are adsorbed on the surface. For surfaces other than mercury the charge per cm.^{2} required to change the potential by one volt was S times 6×10^{-6}. They obtained the following values for S: smooth platinum, 2.0; platinum black, 2000; sandpapered nickel, 10; oxidized and reduced nickel, 50. They interpret this S as the ratio of the true area to apparent area. Their values are considerably greater than those expected from Tonks' theoretical calculations.

As a result of this it is my opinion that a considerable amount of careful work must be done before reliable values of S are obtained for

thermionic cathodes. For the present it would seem best to consider S as an unknown whose value lies somewhere between 1 and 10 for rough surfaces such as those on oxide coated filaments, and between 1 and 2 for relatively smooth surfaces such as tungsten. The exact value will no doubt depend on the exact treatment of each surface.

Fortunately the uncertainty of our knowledge of S does not seriously affect our correlation made above. It is necessary to divide values of i and A in the empirical equations by S to reduce to the basis of true surface area before comparing them with theoretical equations. The observed values of i and A should thus be reduced by 25 to 50 per cent for smooth surfaces and by larger values for rough surfaces. Thus in the case of surfaces, such as tungsten, molybdenum and tantalum, for which A has the value 60, the true A should be between about 30 and 45 as compared with a theoretical value of 120. Since the deviations from 120 are due to a temperature dependence of the work function, it means that we must postulate a somewhat larger value of α in equation (34) than otherwise.

On the Reflection Coefficient

There is still another topic that enters into the correlation of experiment and theory, namely the reflection coefficient. Thus far we have assumed that every electron whose normal component of velocity exceeded a certain value escaped while those having less than this value failed to escape. On the classical viewpoint this assumption is justified but on the quantum-mechanical viewpoint there is a finite probability that the electron considered as a wave will be reflected at the surface even though its velocity is such that it could escape; also a wave electron has a finite probability of passing through a potential peak when classically its velocity is not large enough to permit it to pass over the top of the peak. Consequently we should include an average transmission coefficient $\overline{D}$ in the theoretical emission formula. $\overline{D} = 1 - \bar{r}$ where $\bar{r}$ is the average reflection coefficient. Equation (22) then becomes

$$i = U(1 - \bar{r})T^2 \exp. (- w/T). \tag{38}$$

A number of writers [2, 3] have attempted to explain the deviations between A and U by postulating such values of $\bar{r}$ that $A = U(1 - \bar{r})$. This explanation is possible only for cases for which $A < U$ since $0 < \bar{r} < 1$. Even when $A < U$ the numerical values turn out to be such that the difference between A and U cannot be accounted for by computed probable values of $\bar{r}$. These values of $\bar{r}$ are determined chiefly by the shape of the curve giving the work an electron must do

to get to various distances from the surface. Only when this work-distance curve is postulated to have a high sharp peak within a few atom diameters from the surface, is it possible to deduce values of $\bar{r}$ which are appreciable. Now we have good reasons [8] for believing that no such peaks exist, and that the maximum of the work distance curve occurs at relatively large distances from the surface in a region where the forces on the electron are given by the well-known image law. For the latter type of curve, the computed value of $\bar{r}$ is less than 0.07 which is negligibly small. Nordheim, who first pointed out that the transmission coefficient might differ from unity says: "However, the exact computation taking into account the image force which must necessarily be considered, has shown that such a rounded-off potential curve yields a value of $\bar{D}$ which differs inappreciably from 1.0." * A more complete case showing that the values of the reflection coefficient are negligibly small is given by Becker and Brattain.[8]

The Effect of Accelerating Fields and Retarding Potentials

Thus far we have considered how the emission current and the work function depend on the emitting surface and its temperature; we have implicitly assumed that the current was "saturated" or that every electron which escaped from the surface was collected by the anode. It is, however, well known that the emission current depends also on the applied fields and the applied potentials. In considering the effects of these fields and potentials we shall incidentally obtain an insight into the nature of some of the forces responsible for the work function.

For simplicity consider a large plane cathode and parallel to it a large plane anode. If the temperature of the cathode is high enough to emit a small but appreciable current, $\log i$ will vary with the potential applied to the anode in the manner shown in Fig. 3. In drawing curve 1 in this figure three more simplifying assumptions have been made; namely (1) that the contact potential between cathode and anode is zero; (2) that all portions of the cathode and anode have the same work function, and (3) that space charge effects are negligible. The effect of these assumptions will be considered later.

The curve in Fig. 3 naturally divides itself into two portions: the part to the left of $V_a = 0$ corresponds to retarding potentials while the part to the right of 0 corresponds to accelerating potentials. In the latter region the current is said to be "saturated" although strictly speaking the current is never saturated but increases indef-

* See Section by Nordheim in Müller-Pouillets *Lehrbuch der Physik*,[4] Vol. IV, "Elektrizitat und Magnetismus," Part IV, p. 294. See also footnote 2 on p. 290.

initely as V increases. Obviously, the effect of V on the current is quite different in the two regions and these two regions require different explanations.

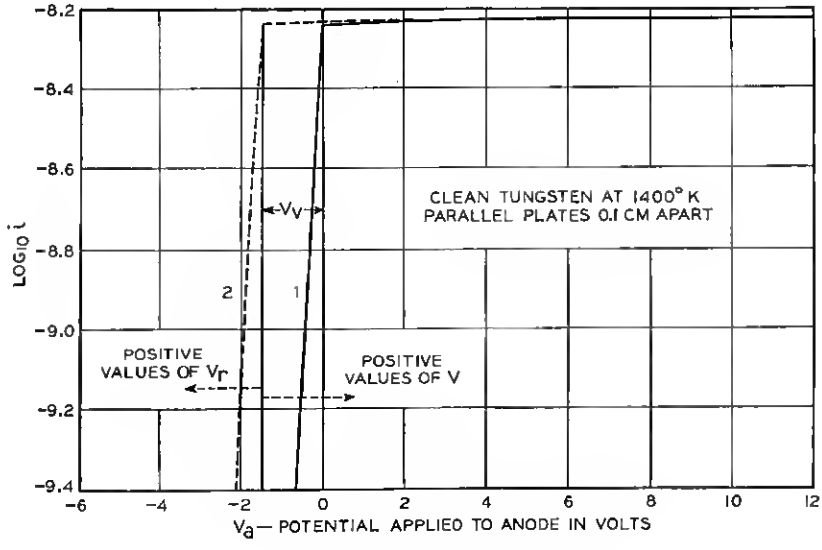


Fig. 3—Retarding potential curves for parallel plates.

Retarding Potentials

Consider first the retarding potential region in which log i decreases linearly with $-V_a$, the applied potential V. The explanation is to be found in extending the theories which were used to derive the Richardson equation. In that derivation it was implicitly assumed that the only forces which the escaping electron had to overcome were the cathode surface forces, and that any electron which escaped from the cathode would reach the anode. If a retarding potential V_r* is applied to the anode then only those electrons whose normal component of velocity u exceeds a value u_a can reach the anode; where u_a is given by

$$mu_a^2/2 = (\varphi_c + V_r)e \tag{39}$$

in the classical case, or

$$mu_a^2/2 = P_m + eV_r \tag{40}$$

in the quantum-mechanical case. Figure 4, curve 1, illustrates the

* In discussing retarding potentials it is convenient to consider retarding potentials as positive even though the anode potential is negative, so that $V_r = -V$.

potential energy of an electron at various distances between the cathode and anode when the anode is V_r volts negative to the cathode. It is tentatively assumed that the anode work function φ_a is the same as the cathode work function φ_c; this is another way of saying that

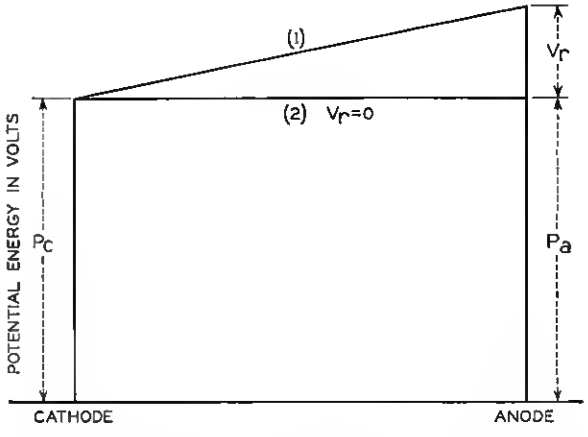


Fig. 4—Potential distribution between parallel plates; $P_a = P_c$.

the contact potential is zero. When $V_r = 0$ nearly all the space between the cathode and anode is field free as shown in curve 2; only in the immediate neighborhood of the cathode or the anode is the electron subjected to any forces. When a retarding potential is applied the electrons must have sufficient energy to pass over the maximum in curve 1, Fig. 4, in order to reach the anode.

To determine the number of electrons that can reach the anode we integrate equation (10) or (20), from $u = u_a$ to $u = \infty$ where u_a is given by equation (39) or (40), respectively. Whether we use the classical or the quantum-mechanical statistics we arrive at the same result.

$$i = Ne = i_0 \exp.\ (-\ V_r e/kT), \tag{41a}$$

or

$$\log i = \log i_0 - (e/2.3kT)V_r, \tag{41b}$$

where $i_0 = i$ when $V_r = 0$. The slope of the straight line in Fig. 3 should thus be $e/2.3kT$.

If φ_a and φ_c are not equal, the field between anode and cathode will not be zero when the applied potential is zero; a contact potential or Volta potential V_V will exist between a point just outside the cathode and a point just outside the anode. To produce zero field a potential

must be applied which neutralizes the Volta potential. The true potential V between anode and cathode is the sum of the applied potential V_a and V_V or

$$V = V_a + V_V. \tag{42}$$

Since

$$V_V = \varphi_c - \varphi_a, \tag{43}$$

$$V = V_a + \varphi_c - \varphi_a. \tag{44}$$

Since

$$V_r = - V,$$

$$V_r = - V_a - (\varphi_c - \varphi_a) = - V_a + \varphi_a - \varphi_c. \tag{45}$$

V_r is the true value of the retarding potential and these values of V_r are to be used in equations (39), (40) and (41). V and V_r are measured from the break point in Fig. 3. Figure 5 illustrates the case

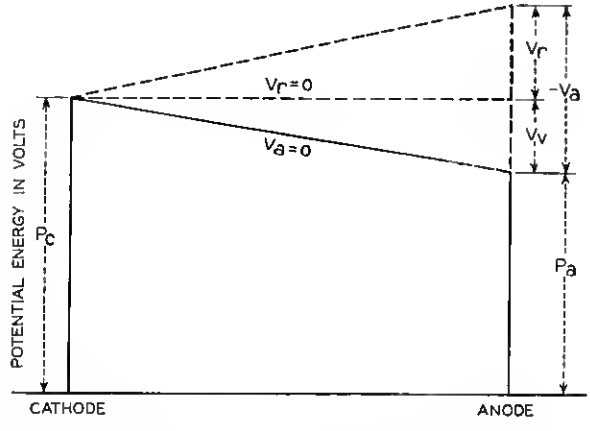


Fig. 5—Potential distribution between parallel plates; $P_a < P_c$.

when $\varphi_c > \varphi_a$. For $V_a = 0$, V is positive and equal to $\varphi_c - \varphi_a$. To produce zero field V_a must be negative and equal to $\varphi_c - \varphi_a$. The dashed line gives the potential energy distribution for a somewhat larger negative applied potential.

When the contact potential is not zero, the break point in the log i *versus* V_a curve will occur when the field is zero or when

$$V_a = - (\varphi_c - \varphi_a).$$

This is illustrated in Fig. 3 by the dashed line for a case for which $\varphi_c > \varphi_a$.

Usually thermionic experiments are not performed with plane parallel cathodes and anodes but with a small cylindrical cathode concentric with a cylindrical anode. In the cylindrical case, the normal or radial component of velocity is not the only one which determines whether the electron will reach the anode. Schottky [19] derived the following formula for this case on the assumption that the emitted electrons leave the filament with a velocity distribution given by Maxwell's law (equation (7)) for a temperature T. As we have seen above both the classical and the Fermi-Dirac theory predict this distribution for the electrons which escape from the filament. This formula replaces equation (41).

$$i = i_0(2/\pi^{\frac{1}{2}}) \left[[V_r e/kT]^{\frac{1}{2}} \exp.\ (-V_r e/kT) + \int_{(V_r e/kT)^{\frac{1}{2}}}^{\infty} \exp.\ (-x^2)dx \right]. \quad (46)$$

It is assumed that the diameter of the cathode is small compared to the diameter of the anode, and that the current is not limited by space charge. Table I gives values of $\log_{10}(i_0/i)$ for values of $V_r e/kT$ taken from an article by Germer.[20]

TABLE I

VALUES OF $\log_{10} i_0/i$ FOR VARIOUS VALUES OF $V_r e/kT$ (GERMER) [20]

$V_r e/kT$	1	2	3	4	5	6
$\log_{10}(i_0/i)$	0.2423	0.5827	0.9523	1.3371	1.7312	2.1318
	7	8	9	10	11	12
	2.5369	2.9455	3.3567	3.7698	4.185	4.6024
	14	16	18	20	25	
	5.4398	6.2812	7.1245	7.9714	10.0978	

Figure 6 shows various ideal plots of log i *versus* V for cathodes of clean tungsten and thoriated tungsten. It is assumed that the anode is clean tungsten. Curves 1, 2 and 3 are for a clean tungsten cathode at temperatures of 1400, 1550 and 1700° K., respectively. Curves 4 and 5 are for a thoriated tungsten cathode at 1400° K. activated to such an extent that the work function is 4.03 and 3.53 volts, respectively. The dashed lines indicate the currents for a plane cathode and parallel anode.

Curves 1, 4 and 5 illustrate an important theorem which follows from the analysis on contact potential given in connection with Figs. 3, 4 and 5. The theorem is: The current collected by an anode is independent of the work function of the cathode provided that the

cathodes are in the same position and have the same temperatures and that the retarding potential is sufficiently great. This theorem was verified experimentally by Davisson.[21]

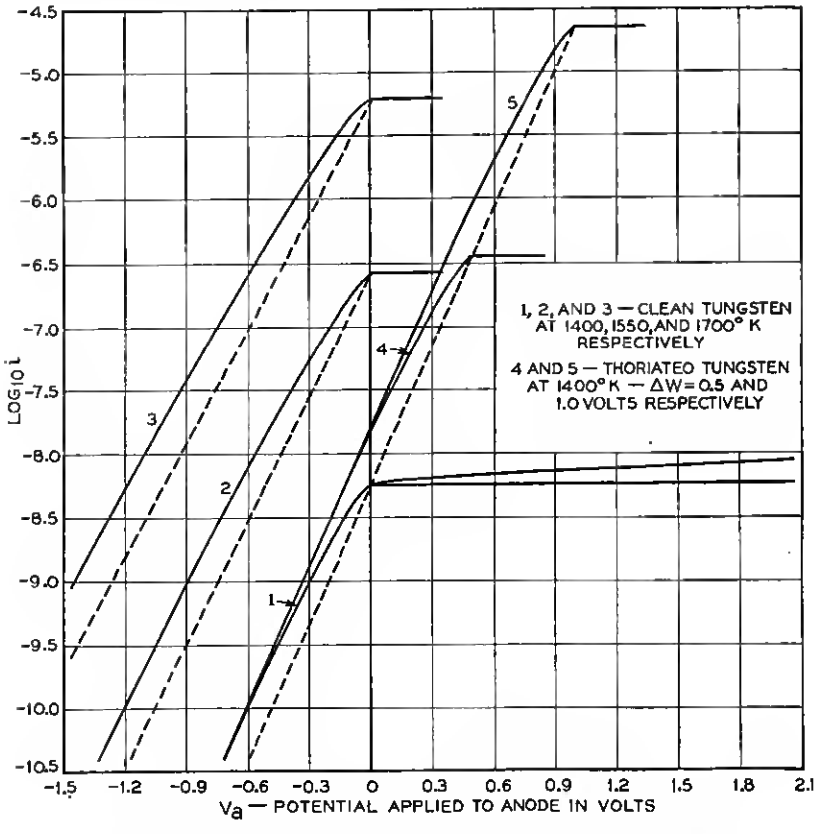


Fig. 6—Retarding potential curves for cylindrical electrodes.

The curves shown in Fig. 6 are for ideal conditions. Experimental conditions frequently differ from ideal conditions in at least five respects: (1) The various portions of the cathode are not at the same potential; (2) the work function of the cathode is non-uniform; (3) the work function of the anode is non-uniform; (4) the temperature of the cathode is non-uniform; and (5) the current is limited by space charge rather than by the applied potential. The last two conditions are discussed in other treatises. The first condition is usually due to the fact that the cathode is a filament and is heated by passing a current through it. As a result the observed curve is the sum of a series of

elementary curves, each one of which is shifted along the V axis by the amount of the potential drop along the filament. Such a sum curve consists of a straight line having the correct slope at sufficiently great retarding potentials; the sharpness of the break point in the curve is, however, destroyed and the slope of the curve for small retarding potentials is decreased, thus simulating the ideal curve for a higher temperature. The best way to obviate this difficulty is to work with equipotential cathodes which are heated indirectly. This makes the construction of the tube more difficult and has been used only by Demski.[22] Most of the work has been done on filaments which were heated intermittently by means of a mechanical or electrical commutator.

In this way Germer,[20] Demski and others have shown that the distribution of thermionically emitted electrons is Maxwellian and corresponds to a temperature which is equal to the temperature of the cathode to within less than 5 per cent. Germer worked with tungsten for a series of temperatures between 1440 and 2475° K. Demski worked with tungsten and with oxide-coated filaments. He used a mechanical and an electrical commutator and also worked with equipotential cathodes. Nottingham [21] and others have reported that for thoriated tungsten and oxide-coated filaments the temperature computed from the shape of the log i *versus* V curve for small retarding potentials was about 1.5 times the temperature of the cathode. Nottingham explains this as due to a sharp peak in the potential distance curve through which a part of the wave electrons can penetrate. In my opinion it is much more likely that these observations are due to non-uniformities in the work function of the cathode and the anode.

If the work function of the cathode is non-uniform, the observed curve should result from the summing up of the currents for a series of curves somewhat similar to curves 1, 4 and 5 in Fig. 6. The sum curve will have the correct slope at sufficiently great retarding potentials; but at low values of V_r the slope should be too small corresponding to too high a temperature. The break point will be less sharp.

If the work function of the anode is non-uniform, the elements of the sum curve will consist of a series of ideal curves shifted parallel to the V axis. The sum curve will again yield correct temperatures at large values of V_r but too high temperatures at small values of V_r. That the work function of cathodes is usually non-uniform will be shown in the next section. It is to be expected that the anode work function will also be non-uniform since the anode is more difficult to heat treat than the cathode. However, when one takes into account the effect of these non-uniformities, it is seen that the experiments

abundantly confirm the theory that the distribution of velocities of thermions is that given by Maxwell's law for an ideal gas.

Accelerating Fields

As illustrated in Fig. 3, when positive potentials are applied to the anode, log i increases continuously; but the rate of increase becomes progressively less so that the current is almost independent of the anode potential. For many purposes one can safely say that the current is saturated; for some purposes, however, it is very important to consider this lack of saturation. More specifically a consideration of this effect gives us direct evidence of some of the forces which are responsible for the work function. Thus, as the electron escapes from the surface, it must overcome certain forces which tend to pull it back. The electrical fields responsible for these forces presumably decrease with the distance from the surface. Call them surface fields F_s. When a positive potential is applied to the anode, a field F_a is produced near the surface of the cathode which tends to help the electrons escape. The value of the field depends on the dimensions of the cathode and anode. This applied field neutralizes the surface field at some distance z from the surface; call this distance the critical distance z_c. If an electron can reach the critical distance, it will escape, since beyond this distance the sum of the applied and surface fields pulls the electron toward the anode. Obviously the critical distance moves closer toward the cathode as the applied field is increased.

A more quantitative concept is obtained by considering the effect of the applied field on the potential energy-distance curve similar to Fig. 4. Now, however, we will be concerned more particularly with regions close to the cathode, so that we will greatly enlarge the distance scale. Figure 7, curve 1, shows such a curve when the true field between cathode and anode is zero. The true field F is the algebraic sum of the applied field F_a and the field produced by the contact potential. Frequently it is convenient to use the term " applied field " in the sense of " true field," i.e., including the contact potential field. An applied field decreases the potential energy of the electron as shown in curve 2. The net potential energy is shown in curve 3.

The maximum height in curves 1 or 3 represents the work function φ in the classical theory or the quantity P_m/e in the quantum theory. In the latter case, since $\varphi e = P_m - K$ from equation (24) and since K does not depend on the applied field,

$$\Delta\varphi = \Delta P_m/e \tag{47}$$

or the decrease in the work function due to an applied field is equal to the decrease in the maximum of the potential energy-distance curve. Since P depends on F, the true field (applied + contact potential field),

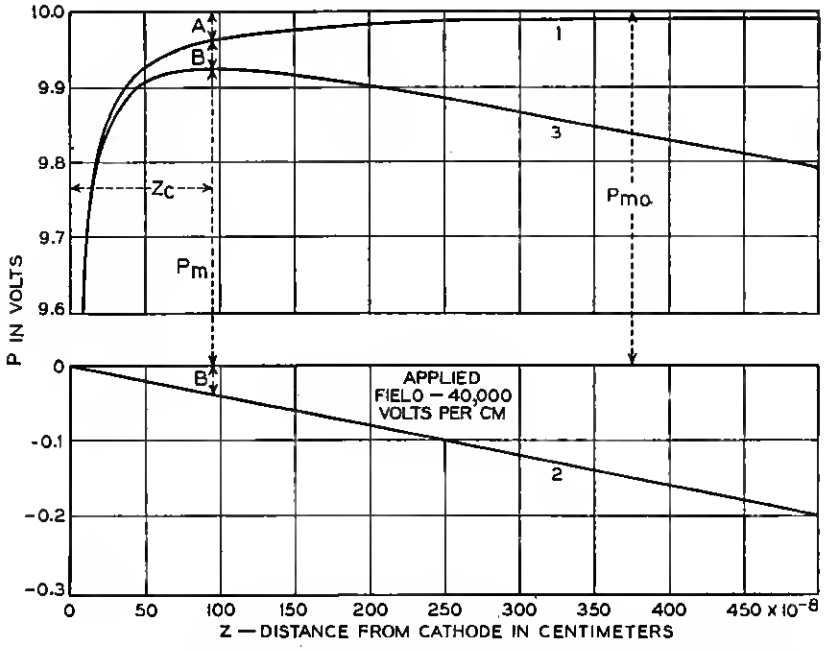


Fig. 7—Potential energy *versus* distance from cathode surface for image force and applied fields of zero or 40,000 volts/cm.

it will be convenient to designate values for curve 1 for which $F = 0$ by the subscript 0. Then

$$P = P_0 - Fez, \tag{48}$$

and

$$dP/dz = dP_0/dz - Fe. \tag{49}$$

At the maximum in the P *versus* z curve, $dP/dz = 0$ and $z = z_c$. Hence

$$(dP_0/dz)\big|_{z_c} = Fe. \tag{50}$$

We require an expression for ΔP_m, the decrease in P_m due to the field F. From Fig. 7 it is clear that

$$\Delta P_m = \text{distance } A + \text{distance } B = P_{m0} - P_0\big|_{z_c} + Fez_c \tag{51}$$

$$d(\Delta P_m)/dF = -\,dP_0/dF\big|_{z=z_c} + Fe(dz_c/dF) + ez_c. \tag{52}$$

Now from equation (50)

$$dP_0/dF\big|_{z=z_c} = dP_0/dz\big|_{z_c} dz_c/dF = F dz_c/dF. \tag{53}$$

Hence from equations (52) and (53)

$$d(\Delta P_m)/dF = ez_c. \tag{54}$$

Combining this with equation (47) we obtain

$$d(\Delta\varphi)/dF = z_c. \tag{55}$$

Now from

$$\log i = \log U - 2 \log T - (\varphi - \Delta\varphi)e/2.3kT = \log i_0 + \Delta\varphi e/2.3kT \tag{56}$$

we obtain

$$d \log i/dF = (d(\Delta\varphi)/dF)e/2.3kT. \tag{57}$$

Combining this with equation (55) we obtain

$$d \log i/dF = (e/2.3kT)z_c. \tag{58}$$

This equation which was first derived by Becker and Mueller [24] allows us to obtain numerical values for z_c from the slope of the experimental log i *versus* F curve. At z_c the surface field F_s is equal to the applied field F. Hence a plot of F_s *versus* z can be obtained, and by integrating this from z to ∞, values of $P_{m0} - P_0$ can be obtained for various values of z greater than some minimum value corresponding to the largest value of F.

A particular case of a surface field, namely, that given by the image law, is especially significant. In this case $F_s = e/4z^2$ and it can be shown that the distances A and B in Fig. 7 are equal. At the critical distance $F = F_s$ and $F = e/4z_c^2$ or

$$z_c = (e/4F)^{\frac{1}{2}}. \tag{59}$$

By substitution in equation (55) and integration from 0 to F it follows that

$$\Delta\varphi = (eF)^{\frac{1}{2}}. \tag{60}$$

Substituting this in equation (56) yields

$$\log i = \log i_0 + (e^{\frac{1}{2}}/2.3kT)\sqrt{F}, = \log i_0 + (1.91/T)\sqrt{F}. \tag{61}$$

This equation, which was first derived by Schottky [35] and is called the Schottky equation or law, predicts that a plot of log i *versus* $\sqrt{F}$ should yield a straight line whose slope is $e^{\frac{1}{2}}/2.3kT$ or $1.91/T$.

Experimental log i *versus* $\sqrt{F}$ plots are found to be straight and to have approximately the right slope for sufficiently high applied fields. At low fields, the line is curved and the experimental slopes are greater than the predicted values. These deviations from Schottky's law are slight in the case of clean surfaces but become quite pronounced for composite surfaces such as thorium on tungsten or cæsium on tungsten. We shall show below that these deviations can be ascribed to non-uniformities in the work function for different regions of the cathode surface. The prediction that the slope should vary as $1/T$ has been verified by Dushman's experiments.[10]

In so far as Schottky's law is verified by experiment, we can conclude that the escaping electron must in certain regions overcome the forces due to its own image and no other forces. Thus for clean surfaces the electron is acted on only by its image force from about 10^{-7} to about 50×10^{-7} cm. from the surface; for composite surfaces this region will depend on the size and degree of the non-uniformities; for a particular surface of thorium on tungsten the image law held from 6×10^{-7} to about 20×10^{-7} cm. When the critical distance is very small, the emission is modified because of sharp points on the surface and because of "intense field" emission.[24] When the critical distance is larger than about 100×10^{-7} or 1×10^{-5} cm. there are apparently other fields superimposed on the image field. These are larger than the image field at these distances and thus cause deviations from the Schottky law. As we shall see later these fields are due to non-uniformities on the surface. From all this we can conclude that an appreciable part of the work function is due to the image force and to other surface fields.

Table II shows values of $\Delta\varphi$, z_c, log i/i_0 and i/i_0 if the surface field is given by the image law.

TABLE II

VALUES OF $\Delta\varphi$, z_c, LOG i/i_0 AND i/i_0 IF THE SURFACE FIELD IS GIVEN BY THE IMAGE LAW

F, volts/cm	0	100	1000	10,000	40,000
$\sqrt{F}$	0	10	31.6	100	200
$\Delta\varphi$, volts	0	0.0038	0.0120	0.0378	0.0755
z_c cm.	∞	1.89×10^{-5}	5.98×10^{-6}	1.89×10^{-6}	9.45×10^{-7}
log i/i_0, $T = 1000°$ K.	0	0.0191	0.0604	0.191	0.382
i/i_0 "	1.000	1.045	1.149	1.553	2.410
log i/i_0, $T = 2000°$ K.	0	0.0096	0.0302	0.096	0.191
i/i_0 "	1.000	1.022	1.072	1.25	1.55

The Use of the Term "Effective Work Function"

There has been a tendency to restrict the term work function to zero field and to use "effective work function" for accelerating

fields.[1, 2, 23] In my opinion this tendency is to be deplored since it is unnecessary and places too much emphasis on zero field. Richardson's equation, and the theories underlying it are just as applicable for accelerating fields as they are to zero field. Since the work function depends on T as well as F, it would be just as logical to coin a new name for the work function at any temperature other than $T = 0$. It seems to me more desirable to retain " work function " in its general sense and to recognize that it may depend on temperature and on the accelerating field. The work function or more precisely the quantity P_m/e would be defined as the work required to take an electron at rest inside the metal to a point at distance z_c from the surface, where z_c is the distance at which the accelerating field is equal to the surface field.

The Effect of Non-uniform Work Functions: Patch Theory

We shall now consider how the emission is altered if the cathode work function is non-uniform. Here again we shall find it necessary to consider the effect of such non-uniformities on the P *vs.* z curves, i.e., on the curves for the potential energy of the electron *versus* distance from the surface. For the present we shall not consider the causes for the mechanism which is responsible for the non-uniformities. We shall assume that the surface work functions are non-uniform. As a consequence, local fields must exist between the various regions having different work functions. The effect of these fields on the log i *vs.* F curve will depend on the size, shape and degree of the non-uniformities.

The Simple Condenser Analog

Consider a simple case: The cathode is uniform except in a circular region of radius R which is covered with a positive charge density σ, a short distance l above the surface. There is induced at a distance l below the surface the image charge density $-\sigma$. These two sheets of charge act like a finite circular condenser. The field between the condenser plates will be $4\pi\sigma$ e.s.u. or $300 \times 4\pi\sigma$ volts/cm. if σ is expressed in e.s.u. If the zero of potential is taken at the surface of the metal or at the center of the condenser, the potential just outside the outer sheet of charge will be $300 \times 4\pi\sigma l$. If the sheet of charge were infinite in extent or if R were several times the distance from cathode to anode, then the field outside the condenser would be zero, and the work function of the patch for electrons would be reduced by $300 \times 4\pi\sigma l$ or by $300 \times 2\pi M$; where $M = 2\sigma l$ the moment per cm.^{2} of surface. Actually there is a field outside the finite condenser which tends to pull an electron back to the surface. The integral of this

field out to infinity or a distance large compared to R is just sufficient to reduce the potential to zero again. Hence when the applied field is zero so that z_c is very large, the work function over the condenser or patch is not reduced at all. Calculations show that if a small accelerating field is applied, the work function is reduced more than it would have been if there had been no condenser. For a sufficiently large applied accelerating field, z_c moves so close to the surface that $z_c \ll R$. At this distance the potential at z_c due to the sheets of charge will not

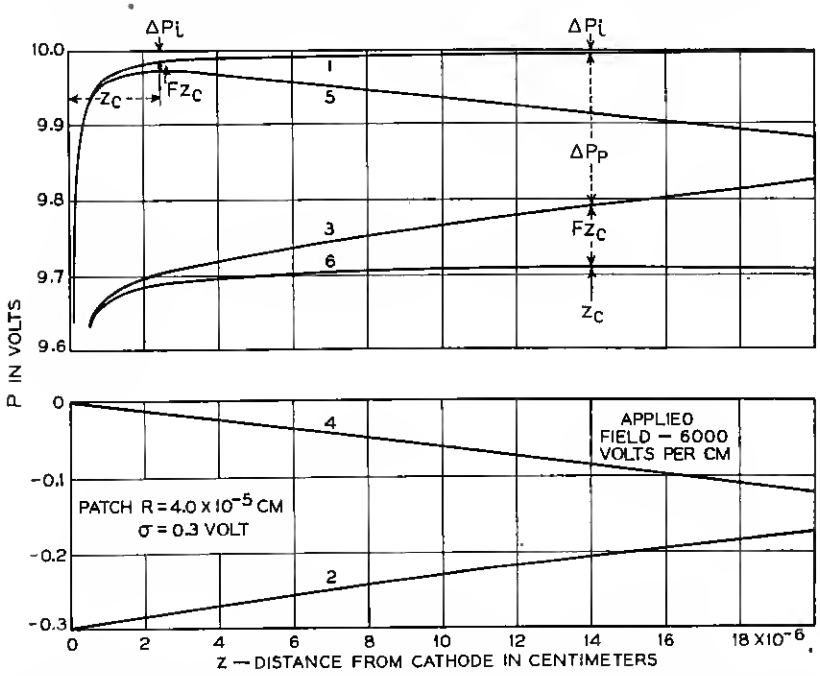


Fig. 8—Potential energy *vs.* distance from cathode surface for electrons moving against image field plus field due to a uniform circular patch for applied fields of zero or 6000 volts/cm.

differ greatly from the potential just outside the condenser. The work function will now be $300 \times 4\pi\sigma l$ less than it would have been without the patch. Hence the extra reduction of the work function due to the patch is zero at zero accelerating field and increases with this field up to its limiting value $300 \times 4\pi\sigma l$.

To treat this case in more detail consider the curves in Fig. 8. Curve 1 is a plot of P_i/e in volts *vs.* z in cm. P_i is the potential energy due to the image force and is given by

$$P_i/e = P_\infty/e - 300e/4z, \tag{62}$$

where P_∞ is the potential energy when $z = \infty$ and $F = 0$; $e = 4.774 \times 10^{-10}$. Curve 2 is the potential energy P_p due to the patch or condenser along a line normal to the surface at the center of the condenser. R has been taken as 4×10^{-5} cm. The equation of this curve is

$$P_p = 1200\pi\sigma l[z/(z^2 + R^2)^{\frac{1}{2}}]. \tag{63}$$

In the derivation of this formula it has been assumed that either $R \gg l$ or else that $z \gg l$; for our case the first of these assumptions will always be fulfilled so that the formula is applicable when z is equal to or larger than l; it is not applicable for z less than l.

Curve 3 is the algebraic sum of P values for curves 1 and 2. It represents the potential energy along the central normal due to the image and patch fields. Curve 4 represents the potential energy due to an applied field of 6000 volts/cm. Curve 5 is the sum of curves 1 and 4; curve 6 that of 3 and 4.

The effect of the applied field is to reduce the critical distance z_c and the work function. A given applied field will reduce z_c more for a clean surface than for one with the patch; but the converse is true for the work function. The reduction in the work function is equal to the reduction in the value of P_m/e. This consists of three parts as indicated in the figure for curve 6. $\Delta P_i/e$ is the decrease in P due to the image forces from $z = z_c$ to $z = \infty$; $\Delta P_p/e$ is the decrease due to the patch field from $z = z_c$ to $z = \infty$; Fz_c is the decrease in P due to the applied field from $z = 0$ to $z = z_c$. These quantities can be evaluated after one has determined the value of z_c as follows:

The peak or maximum in curve 6 occurs at a value of $z = z_c$ at which

$$dP/dz = dP_i/dz + dP_p/dz = Fe. \tag{64}$$

From equation (62) $dP_i/dz = 3.58 \times 10^{-8}/z^2$,

and from equation (63)

$$\frac{dP_p}{dz} = -(1200\pi\sigma l)e\left[-\frac{1}{(z^2 + R^2)^{\frac{1}{2}}} + \frac{z^2}{(z^2 + R^2)^{\frac{3}{2}}}\right]$$

$$= 1200\pi\sigma le\frac{R^2}{(z^2 + R^2)^{\frac{3}{2}}}$$

so that

$$F = \frac{3.58 \times 10^{-8}}{z^2} + 1200\pi\sigma l\left[\frac{R^2}{(z^2 + R^2)^{\frac{3}{2}}}\right]. \tag{65}$$

From this equation F is plotted for various values of z. For any value of F a value of z can be read off. This value of z will be the critical

distance z_c. To obtain $\Delta\varphi$, z_c is substituted in the equation,

$$e\Delta\varphi = \Delta P_i + \Delta P_p + Fz_c = P_\infty - \frac{3.58 \times 10^{-8}}{z_c} + 1200\pi\sigma l\left(1 - \frac{z_c}{(z^2 + R^2)^{\frac{1}{2}}} + Fz_c\right). \quad (66)$$

This $\Delta\varphi$ is the decrease in the work function for a region near the center of the patch. For other regions on the patch $\Delta\varphi$ will be smaller; for regions on the uncovered portion of the surface $\Delta\varphi$ will be still smaller until at large distances from the patch $\Delta\varphi$ will correspond to the $\Delta\varphi$ appropriate for the image law.

To obtain the effect of the patch on the log i *vs.* F or log i *vs.* $\sqrt{F}$ curve it is necessary to divide the entire surface into small regions, compute $\Delta\varphi$ for each and substitute these in equation (56); the values of i_0 in this equation are the same for all regions of equal area since at large distances P_∞ has the same value over all regions. The values of i are then added up for all regions and log i is plotted *vs.* $\sqrt{F}$. Since this process is very tedious, and since in most thermionic experiments one is not likely to deal with a single patch, it is not worth while to make such an exact computation. It is, however, instructive to make some further computations based on simplifying assumptions.

Suppose we assume: (1) That for all regions on the patch, $\Delta\varphi$ has the same value as for the central region, and (2) that the current from the patch is large compared to the current from the uncovered portions of the surface. These assumptions approximate the true conditions for some cases and the errors due to the first assumption tend to balance out those due to the second. If "a" is the area of the patch then

$$\log ai = \log a + \log U + 2\log T - \varphi e/2.3kT + (\Delta\varphi)e/2.3kT = \log ai_0 + \Delta\varphi e/2.3kT, \quad (67)$$

where ai_0 is the current from the patch area when $F = 0$.

Hence
$$\log i/i_0 = (\Delta\varphi e/2.3kT). \quad (68)$$

Values of $\Delta\varphi e$ obtained from equation (66) are substituted in equation (68) and log i/i_0 is plotted as a function of $\sqrt{F}$. Figures 9 and 10 show such plots.

Figure 9 shows the effect of varying the radius R of the patch while the charge density σ is kept constant. The value of σ is so chosen that

$1200\pi\sigma l$ is equal to 0.3 volt. It will be convenient to treat σ as if it were expressed in volts, i.e., as if σ stood for $1200\pi\sigma l$. If the patch were very large σ in volts would be the decrease in the work function due to the patch. It is to be noted that a typical curve starts along a

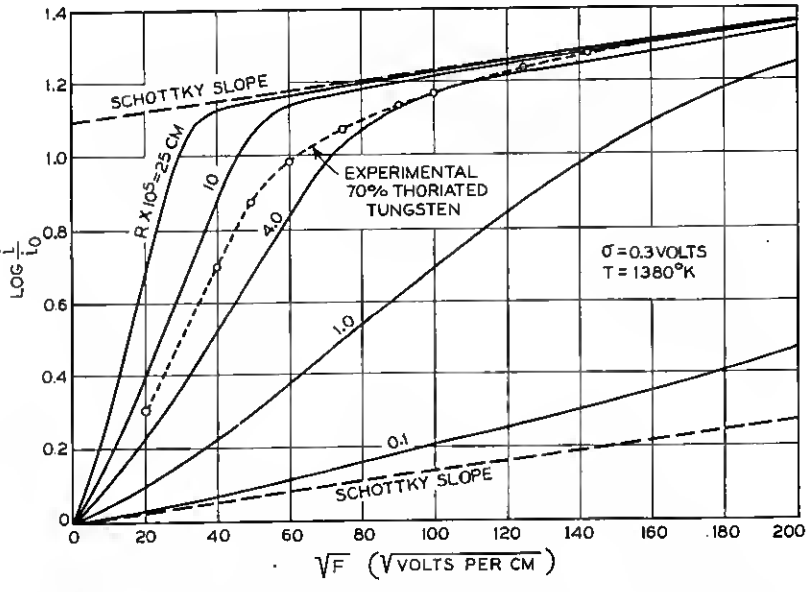


Fig. 9—Variation of emission current with applied field for circular patches of various sizes; T and σ constant. Comparison with experimental curve for thoriated tungsten.

line having the Schottky slope; but soon it rises at a much more rapid rate and continues until it almost reaches an upper line having the Schottky slope; then it bends rather sharply and approaches this line asymptotically. For the larger patches, the curve starts to rise at very small values of $\sqrt{F}$ and it is very steep. For smaller values of R, the curve follows the lower Schottky line for an appreciable distance and its slope never attains very large values; also the place at which it bends toward the upper Schottky line moves to large values of $\sqrt{F}$. Note also that as long as σ is constant all curves are bounded by the same two Schottky lines.

Figure 10 shows the effect of varying σ while R is kept constant. The distance between an upper Schottky line and the lower Schottky line is directly proportional to σ. In fact this shift is given by

$$\Delta \log i/i_0 = \sigma e/2.3kT. \tag{69}$$

It is also apparent from Fig. 10 that increasing σ results in a steeper curve and in an increase in the value of $\sqrt{F}$ at which the curve bends toward the upper Schottky line.

Actually, of course, the observed current will be composed of the current from the uniform part as well as that from the patch. The

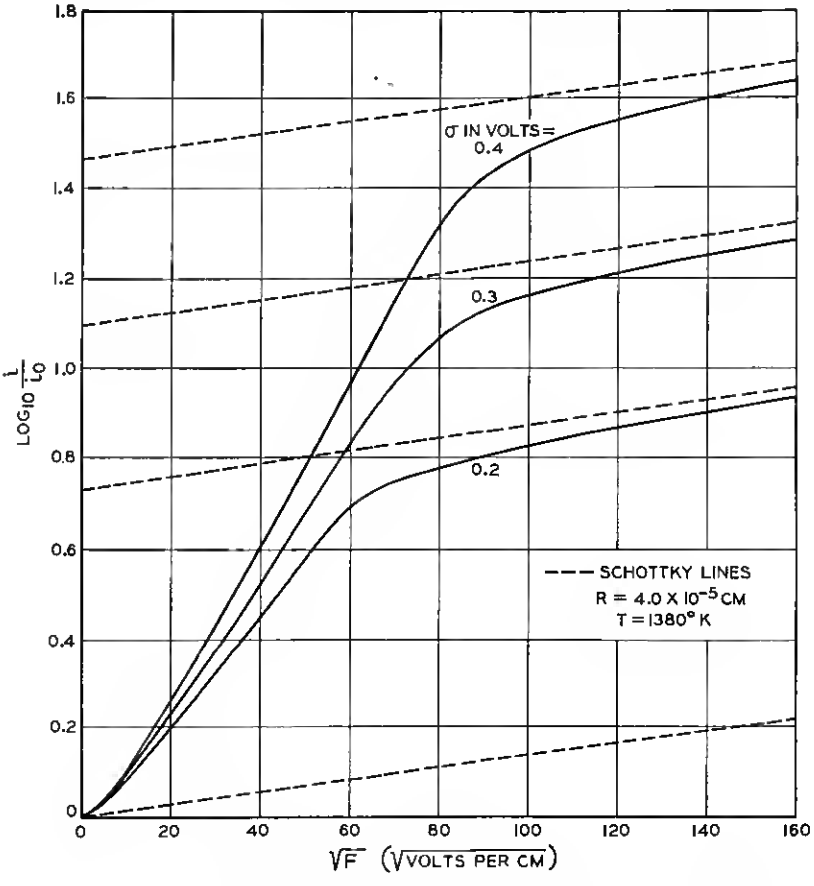


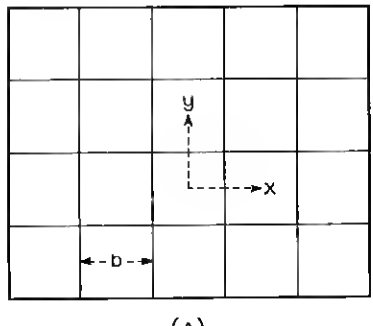
Fig. 10—Variation of emission current with applied field for circular patches; σ variable, T and R constant.

amount by which the patch current influences the total current will depend on the ratio of the patch area to the total area.

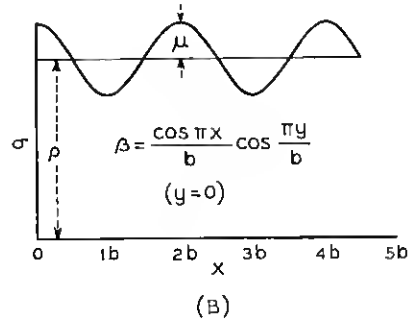
The Hill and Valley Checkerboard

Instead of being covered with a single patch, the surface in the case of most thermionic cathodes consists of numerous patches of varying

sizes and varying work functions above and below some mean value. To treat this case would obviously require very complex expressions. We can simplify the problem without departing too far from actual conditions by postulating a surface which is divided up into a large number of squares arranged in a checkerboard fashion. We might suppose that all black squares have the same σ and all white squares are bare or else have a smaller σ. It turns out, however, that the formulas and the computations are much simpler if we suppose σ is largest at the center of each black square and is least at the center of each white square; between the centers σ is given by a cosine law. In other words on the black squares we have a hill of charge while on the white squares we have a valley of charge. It will be found that such a charge distribution predicts changes in emission with applied fields, which agree rather well with experiment if the size of the squares is comparable to the crystal size and the difference in contact potential between the hills and valleys corresponds to several tenths of a volt.



(A)



(B)

C	B	C	C'	B'	C'
B	A	B	B'	A'	B'
C	B	C	C'	B'	C'

SUB CHECKERS

(C)

Fig. 11—A. Checkerboard array. B. Charge distribution for hill and valley checkerboard. C. Subdivision of checkers.

To represent such a charge distribution, choose the origin of coordinates at the center of a covered square; let x be measured parallel to one edge of the squares while y is measured perpendicular to this edge as indicated in Fig. 11A. Let the length of each square be b.

Then the surface charge density σ is given by

$$\sigma = \rho + \mu \cos(\pi x/b) \cos(\pi y/b) = \rho + \mu\beta, \qquad (70)$$

in which ρ is the mean value of σ, $\rho + \mu$ is the maximum value of σ, $\rho - \mu$ is the minimum σ, and $\beta = \cos(\pi x/b) \cos(\pi y/b)$; β has values between $+1$ and -1. It readily follows that along the edges of the squares $\beta = 0$ and $\sigma = \rho$. Figure 11B shows σ as a function of x when $y = 0$, b, $2b$ or nb.

The great advantage of this particular charge distribution is that we can represent the potential due to this charge and its image *at any point* above the surface by means of a comparatively simple formula, *viz.*,

$$P_\sigma/e = -300 \times 4\pi l[\rho + \mu\beta \exp(-\sqrt{2}\pi z/b)], \qquad (71)^*$$

P_σ is the potential energy of an electron due to the charge distribution at a point which is z cm. above the surface over a region at which the charge density is σ. The charge distribution is located in a plane which is l cm. above the surface. This charge distribution induces a corresponding negative charge distribution at $z = -l$, i.e., l cm. below the surface. ρ and μ are in e.s.u. of charge per cm.^{2} Sometimes it will be convenient to treat ρ and μ as if they were expressed in volts, i.e., as if ρ and μ stood for $1200\pi\rho l$ or $1200\pi\mu l$, respectively. The total potential energy of an electron at z cm. from the surface is given by P in

$$\begin{aligned} P &= P_i + P_\sigma - Fez \\ &= P_\infty - 300e/4z - 1200\pi l[\rho + \mu\beta \exp(-z\sqrt{2}\pi/b)] - Fez, \end{aligned} \qquad (72)$$

where $P_i = P_\infty - 300e/4z$ and $e = 4.774 \times 10^{-10}$.

In Fig. 12, curve 1 shows P_i *vs.* z; curve 2 shows P_σ for various values of β; the curve for $\beta = +1$ is for the normal taken at the center of a hill; $\beta = -1$ is for the center of a valley; $\beta = 0$ is for the edge of the squares; all other curves must lie between those for $\beta = +1$ and $\beta = -1$. Curves 3 show $P_i + P_\sigma$ for $\beta = 1$, 0 and -1.

For all values of β between 1 and 0 the curves have the same maximum value which occurs when $z_c = \infty$. The value of this maximum is $P_\infty - 1200\pi l\rho$. This means that for all points of a hill checker the work function is reduced by the same amount, namely, $1200\pi l\rho$;

* For the derivation of this and several other formulas I am indebted to Professor V. Rojansky now at Union College, who worked with me on this problem in the summer of 1930.

and this amount is the same as would occur if the charge density of the hill and valley checkers were uniformly distributed over the entire surface. On the other hand for β between 0 and -1, i.e., for points

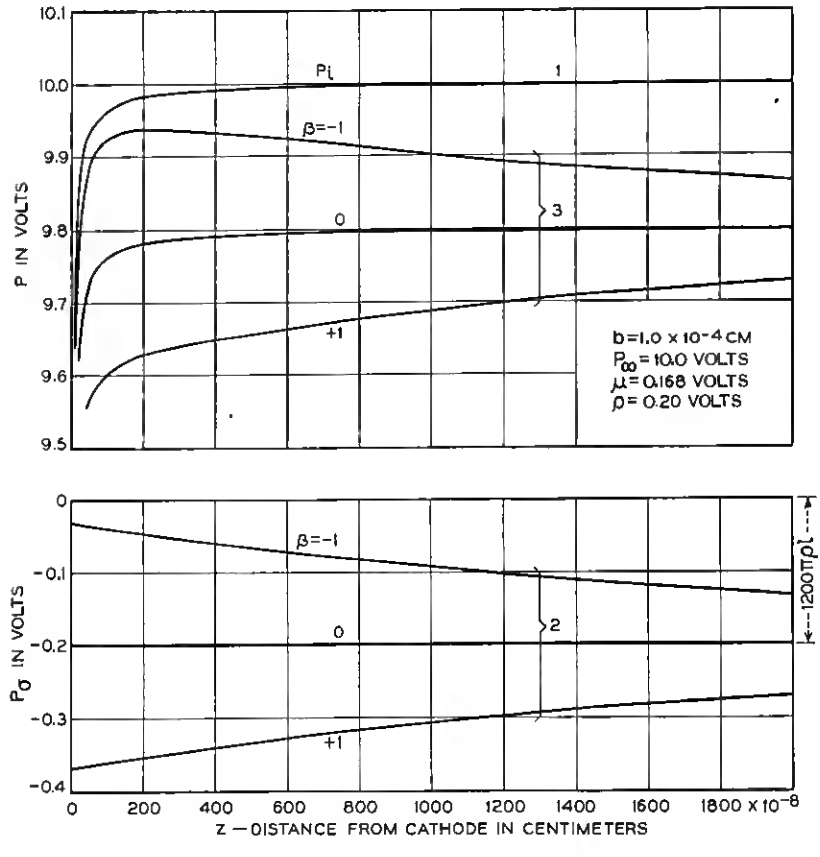


Fig. 12—Potential energy *vs.* distance from the surface above the center of the different subcheckers of the hill and valley checkerboard; zero applied field.

on a valley checker, the value of z_c depends on the particular point chosen. The reduction in the work function is less than $1200\pi l\rho$ and varies from point to point.

The next step is to ascertain the effect of an accelerating field on these P *vs.* z curves. Figure 13 shows this for $\beta = 1, \frac{1}{2}, \frac{1}{4}, -\frac{1}{4}, -\frac{1}{2}$ and -1, respectively, for $F = 5000$ volts/cm. The effect of the field is to decrease z_c and P_m. From Figs. 12 and 13 it follows that the decrease in z_c and in P_m due to F varies with β and is much larger for

points above a hill checker than for points above a valley checker. Values of z_c and P_m are shown in Fig. 14.

Figure 14A shows z_c at various values of x for $y = 0$; it also shows z_c at various values of x for $y = b/3$. Figure 14B shows P_m for these

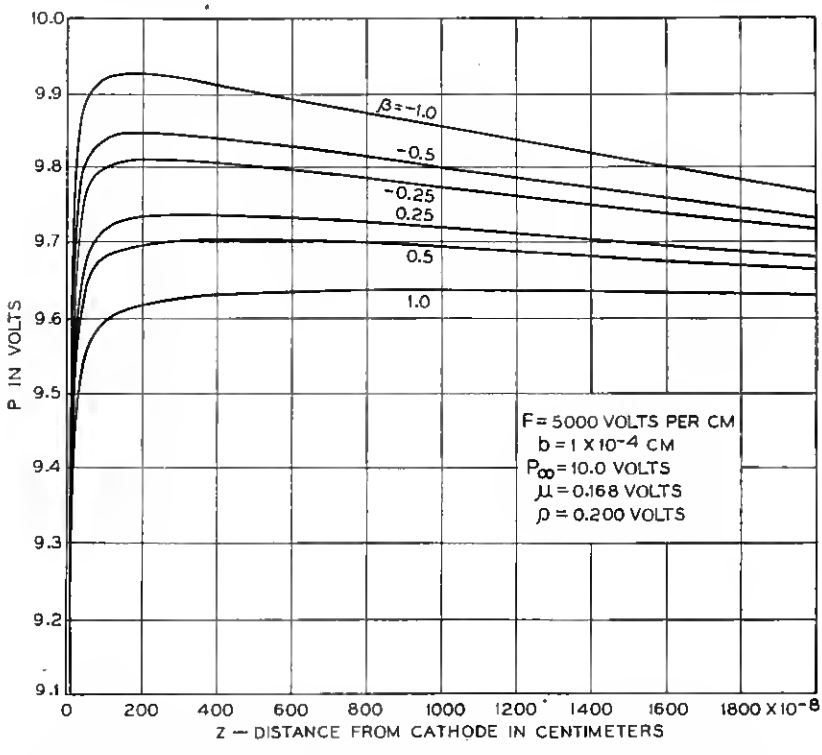


Fig. 13—Potential energy *vs.* distance from the surface above the center of the different subcheckers of hill and valley checkerboard; applied field of 5000 volts/cm.

same values of x and y. Since $P_m - K = \varphi e$, it is clear from these figures that different portions of the surface will have different work functions, or stated more precisely, the energy an electron must have to cross the critical surface (loci of the values of z_c) depends upon where it crosses the critical surface. This in turn means that the chance that a given electron will escape depends not only on its normal component of velocity but also on the place at which it leaves the surface and on the angle its path makes with the surface.

To compute accurately the emission current is a very difficult task. It would appear that the following procedure should give a good approximation to the true current. Divide a "hill" square and a

neighboring "valley" square into nine subsquares each, as indicated in Fig. 11C. It is apparent that the B squares are all alike; similarly the C, B' and C' squares are alike. Determine the β for the center of each subsquare. For the A, B, C, A', B' and C' subsquares the values of β

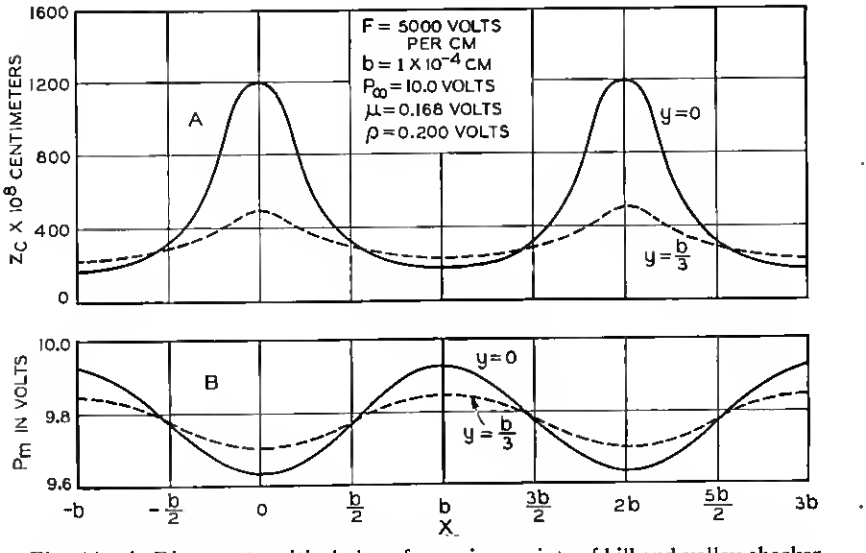


Fig. 14—A. Distance to critical plane for various points of hill and valley checkerboard. B. Potential energy at critical plane for various points of hill and valley checkerboard.

are, respectively, 1, $\frac{1}{2}$, $\frac{1}{4}$, -1, $-\frac{1}{2}$ and $-\frac{1}{4}$. Then compute the P *vs.* z curve for the normals at the center of each subsquare. These are the curves shown in Fig. 13. Next compute the current for each subsquare assuming this to be the same as it would be for an equal area of a large surface having the same work function as that for the center of the subsquare. This is equivalent to assuming that the effect of the velocity components in the x and y directions average out. Do this for various values of F. At each F, add up the currents for all 18 subsquares and multiply this by $1/2b^2$, the number of pairs of squares in a cm.^{2} This will give the current per cm.^{2} of surface for various values of F.

Figure 15 shows the values of the current for the various subsquares; more precisely it shows $\log i/i_{ao}$ *vs.* $\sqrt{F}$ where i_{ao} is the current that would be obtained from an equal area at zero field for a surface covered with a charge density ρ which is the average value of the charge density for the entire surface. i_{ao} is equal to the current from the hill or active subsquares at zero field, but the current from the valley

subsquares is usually less than i_{ao}. The reason for this is clear from an inspection of the curves in Fig. 12. Figure 15 also shows log i/i_{ao} for the average current for 18 subsquares in a pair of hill and valley squares; more precisely, i/i_{ao} is the sum of the current for the 18 subsquares divided by the current that would be obtained if the charge density

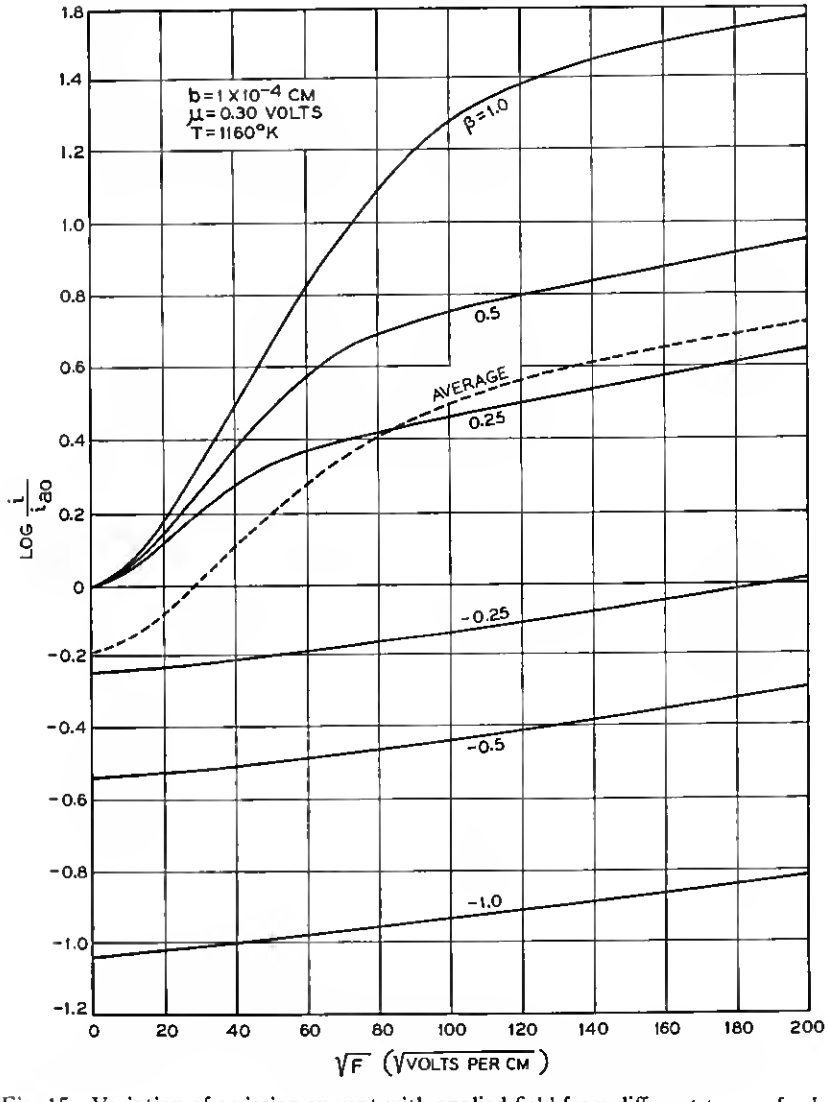


Fig. 15—Variation of emission current with applied field from different types of subcheckers for hill and valley checkerboard; also average curve.

were uniform and the applied field were zero. As the field is increased the *relative* contribution to the sum current from the central hill square becomes larger and larger, while that from the valley squares becomes less and less. At very large fields the curve for any subsquare approaches a straight line whose slope is the Schottky slope; hence the sum curve also approaches a straight line having this same slope. The area which is now contributing most of the current is, however, considerably less than the entire area. Roughly speaking, one might say that at low fields something more than half the area is "effective" in emitting electrons; as the field increases the "effective" area decreases; at large fields and high values of μ, 50 per cent of the total current comes from about 5 per cent of the total surface; one might say that the "effective" area is approximately twice as large as this or 10 per cent. The values of the "effective" areas depend on the value of μ: as μ increases the "effective" area decreases.

Such average curves as the one shown in Fig. 15 depend on three variables, b, μ and T. This dependence is illustrated in Figs. 16, 17 and 18; in each case two of the variables are kept constant. Figure 16 shows log i/i_{ao} *vs.* $\sqrt{F}$ for three values of b. This curve is similar to Fig. 9 for a single circular patch. All the curves still approach a Schottky line at high values of $\sqrt{F}$ but because of the averaging process they do not start out from a common value when $F = 0$ and the initial slope is not equal to the Schottky slope. In both figures as the size of the patch decreases, the curves get less steep and the place at which the curves bend over toward the upper Schottky line moves to higher values of F. Note also that beyond this bend, the curves are approximately straight but only approximately and that the values are still somewhat below the theoretical Schottky line.

This theoretical line has an intercept given by

$$\log i_S/i_{ao} = \log (1/18)[\exp \mu e/kT + \exp - \mu e/kT + 4(\exp \mu e/2kT + \exp - \mu e/2kT) + 4(\exp \mu e/4kT + \exp - \mu e/4kT)]. \quad (73)$$

This equation which defines i_S is based on a subdivision of the hill and valley checker into 18 subcheckers. The exponentials contain the value of β for each type of subchecker, in this case 1, $\frac{1}{2}$ and $\frac{1}{4}$. If each checker were divided into a larger number of subcheckers the number of terms would be increased, but fortunately the value of log i_S/i_{ao} would not be greatly affected. This is particularly true as long as $\mu e/kT$ is less than 5. Since $e/kT \sim 0.1$ this means that values of log i_S/i_{ao} are essentially correct for μ less than 0.5 volt. We have plotted log i_S/i_{ao}

vs. $\mu e/kT$ for 9 subcheckers and for 25 subcheckers. For $\mu e/kT = 5$ the former curve is only 4 per cent higher than the latter; for $\mu e/kT = 10$ the difference is about 6 per cent; for $\mu e/kT$ less than 5 the difference is negligible. This makes us feel that our average curves which are

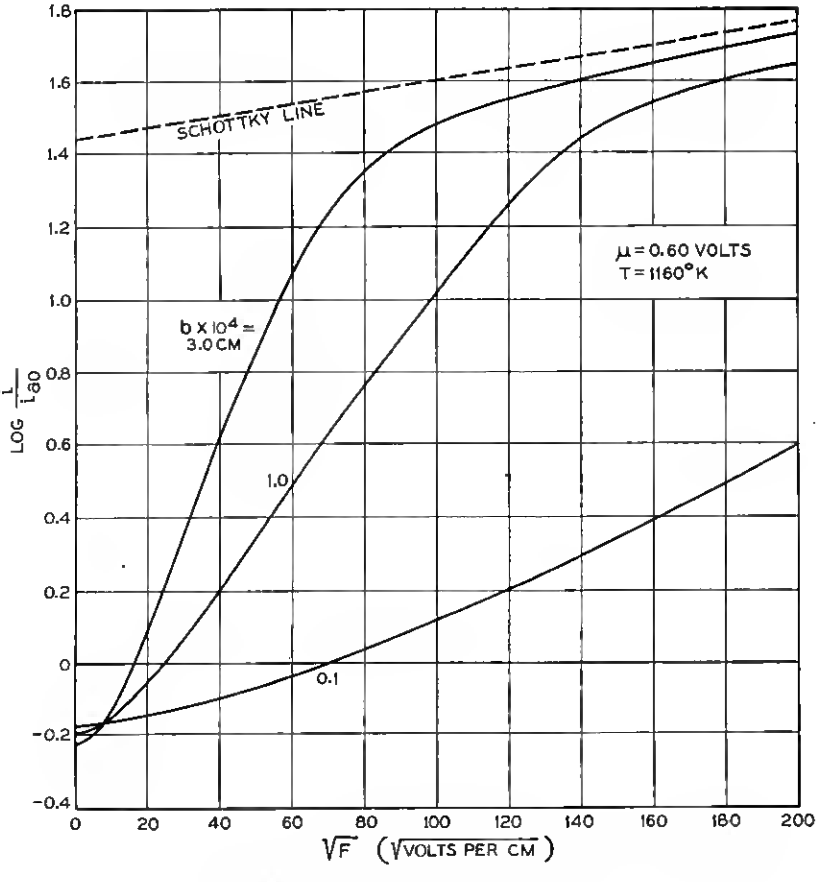


Fig. 16—Variation of average emission current with applied field for hill and valley checkerboard; b variable, T and μ fixed.

based on 18 subcheckers are essentially the same as would be obtained for a much larger number of subcheckers. Equation (73) is analogous to equation (69). Note that it depends on μ but not on b.

Figure 16 might have shown log i/i_S instead of log i/i_{ao}. To convert it to this coordinate it is merely necessary to reduce all values of the ordinate by log i_S/i_{ao} or by 1.44. The advantages of this ordinate will become apparent when we compare theoretical and experimental

curves. For the latter it is quite easy to obtain log i_S but not so easy to determine log i_{ao}.

In Fig. 17 we have thus shown log i/i_S *vs.* $\sqrt{F}$ for constant b and T but varying μ. Had log i/i_{ao} been plotted the curves would have been

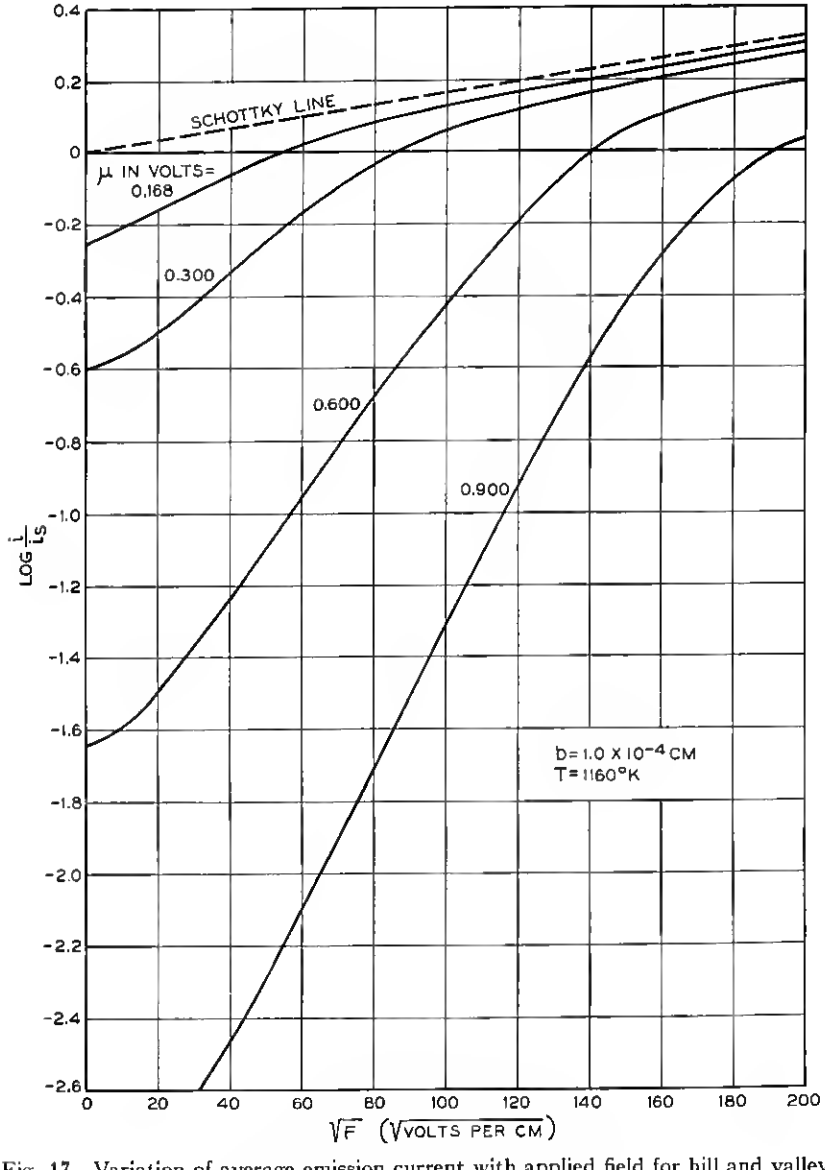


Fig. 17—Variation of average emission current with applied field for hill and valley checkerboard; μ variable, T and b fixed.

close together at $F = 0$ but would have approached upper Schottky lines whose position varied greatly. As it is, all curves approach the same upper Schottky line but "fan out" toward lower values of $\sqrt{F}$. Note that the curves do not cross over as they do in Fig. 16; note also that as μ increases the curves become steeper and the bend toward the Schottky line occurs at larger values of $\sqrt{F}$.

At first sight it might appear that by plotting $T \log i/i_S$ it would be possible to eliminate T as a parameter; while this is true for any one subchecker for which β is a constant, it is not true for the average curves. This is illustrated in Fig. 18 which shows $(kT/e) \log i/i_S$ *vs.*

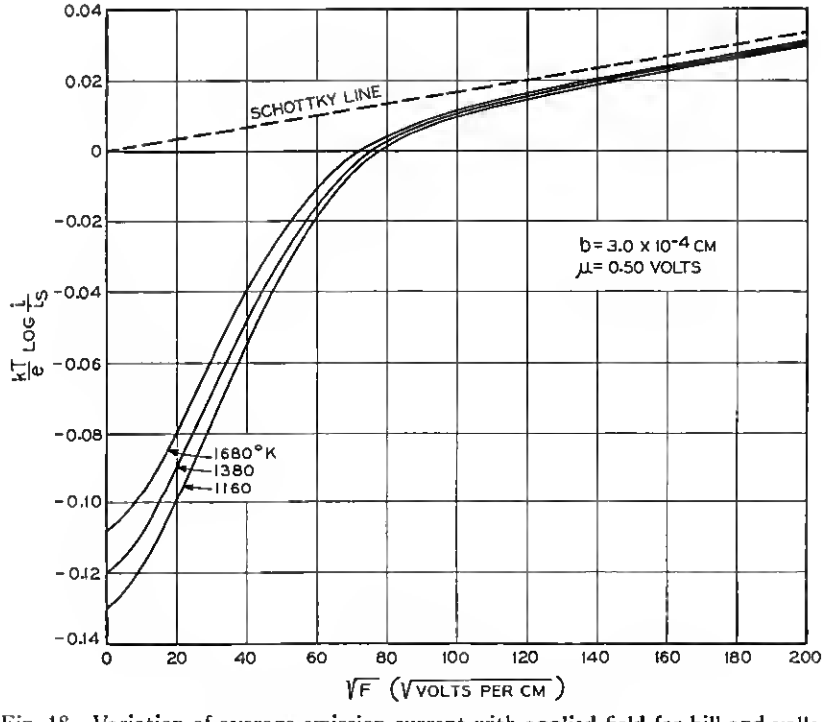


Fig. 18—Variation of average emission current with applied field for hill and valley checkerboard; T variable, b and μ fixed.

$\sqrt{F}$ for constant b and μ and varying T. Note that the departure from the Schottky law is more pronounced for the low temperatures.

Comparison Between Theory and Experiment

It has been found possible to choose values of b and μ such that the calculated average curve fits a given experimental curve over its entire range. The agreement is not perfect; but a perfect fit is not to be

expected when one considers that in an experimental filament the patches are not all of the same size and neighboring patches do not have the same differences in work function. One example is illustrated in Fig. 19 which shows $(kT/e) \log i/i_{ao}$ *vs.* $\sqrt{F}$ for a thoriated tungsten

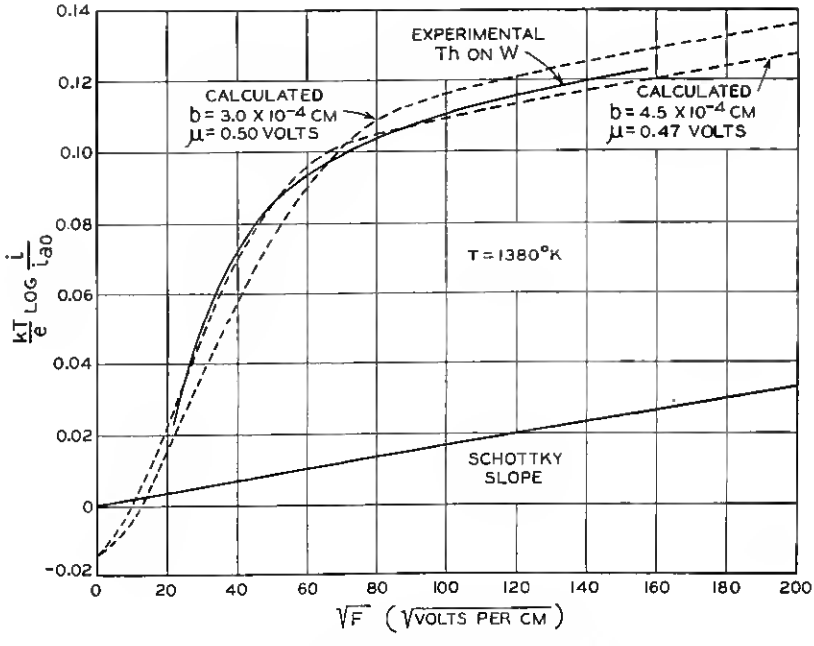


Fig. 19—Comparison of experimental and calculated curves.

filament, 70 per cent of whose surface was covered with thorium. For this curve i_{ao} is merely an arbitrary constant, chosen so as to give the best fit with the computed curves. By comparing the experimental curve with calculated curves in Figs. 16 and 17, we decided to try $b = 3.0 \times 10^{-4}$ cm. and $\mu = 0.50$ volt.* The average curve for these values is shown in the figure. It appeared that b was too small and μ was too large and a new curve was computed and plotted assuming $b = 4.5 \times 10^{-4}$ cm. and $\mu = 0.47$ volt. The agreement is better. Probably a slightly better fit could be obtained with $b = 4.0 \times 10^{-4}$ cm. and $\mu = 0.48$ volt. For other thoriated tungsten filaments we have found values of b from 1×10^{-4} to 1×10^{-3} and μ values from 0.25 to 0.48 volt.

* In choosing these values some allowance had to be made for the fact that the experimental curve was for $T = 1380°$ K., while the calculated curves were for $T = 1160°$ K.

An examination of a number of thoriated tungsten filaments with a microscope showed that the diameter of the tungsten crystals was of the same order as the values of b given above, namely 10^{-4} to 10^{-3} cm. These filaments had been given the customary heat treatment in a vacuum at temperatures near 2800° K. for times measured in minutes and at temperatures near 2100° K. for many hours. It was natural, therefore, to form the hypothesis that different crystals have different adsorptive properties and that consequently different crystals in the same filament should be covered with varying amounts of thorium. Different crystals will then have different work functions. The values of μ found by the above analysis are consistent with this hypothesis since 2μ, which is the difference in work function between a hill checker and a valley checker, is always considerably less than 2.0 volts which is the maximum difference in work function between clean tungsten and thorium on tungsten.

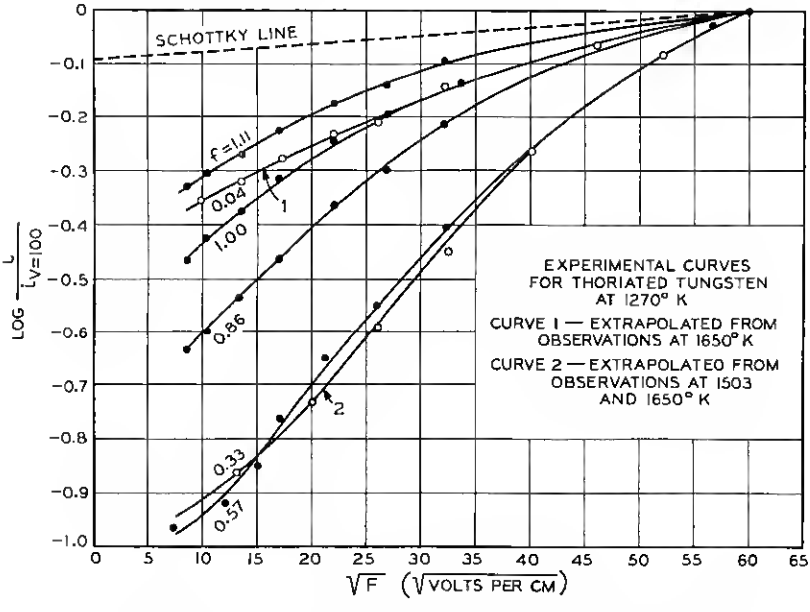


Fig. 20—Experimental log i *vs.* $\sqrt{F}$ curves for thoriated tungsten for various values of f at 1270° K. f = fraction of surface covered with thorium.

If this hypothesis is true, then as a thoriated tungsten filament is activated, the curves for the various stages of activation should all correspond to approximately the same value of b. Figure 20 shows such a family of experimental curves taken by W. H. Brattain of these laboratories. Curves 3, 4, 5 and 6 were taken at $T = 1270°$ K.; curve

1 is extrapolated from data at $T = 1650°$ K.; curve 2 is extrapolated from data at 1503 and 1650° K. The extrapolations were made by extending Richardson lines; curves 1 and 2 are, of course, not quite as certain as data taken at 1270° K. The currents at any V or F vary by large factors as f, the fraction of the surface covered with thorium varies. In order to compare the curves more effectively log $i/i_{V=100}$ has been plotted. This is equivalent to shifting the log i curves until they pass through a common point at $V = 100$ or $\sqrt{F} = 60.5$. A comparison of this family of curves with the computed curves in Figs. 16 and 17 shows a great similarity with Fig. 17 but not with Fig. 16. In Fig. 17, b was constant while μ was varied. From this similarity it follows that the experimental curves in Fig. 20 are consistent with an approximately constant value of b but varying μ.

From the position and shapes of the curves in Fig. 20, we estimate that the value of b or the crystal size is about 4×10^{-4} cm. This value is probably too large for curve 1 and too small for curve 6 but unless a complete analysis were made, it is not desirable to discuss small variations in b. It is apparent from the figure that μ changes with f. We have estimated the following values of μ, the second significant figure being in doubt:

TABLE III

VALUES OF f, THE FRACTION OF SURFACE COVERED WITH THORIUM AND VALUES OF μ IN VOLTS (EQ. 71) μ IN VOLTS = $1200\pi\mu l$.

f =	0.04	0.33	0.57	0.86	1.0	1.11
μ(volts) =	0.23	0.44	0.45	0.36	0.28	0.23

These values of μ are reasonable. Furthermore, the way in which μ varies with f is to be expected from the shape of the work function *vs.* f curve which will be discussed later under adsorption.

A particularly interesting test of the patch theory is furnished by Taylor and Langmuir's [25] electron emission from cesium on tungsten because in this case the crystal size of the tungsten is known. In Figs. 11, 12 and 13 of their article they give log i *vs.* V or $\sqrt{V}$ curves. Since the diameter of the filament is given as 2 mils, it is possible to convert values of V to values of F and to obtain log i *vs.* $\sqrt{F}$ curves. We have done this for the curve for $\theta = 0.60$ and have then analyzed it on the basis of the hill and valley theory. This analysis gave $b = 0.8 \times 10^{-3}$ cm. and $\mu = 0.20$ volt. The article states * "the average grain size in these filaments was about one-fifth the diameter of the wire." So that the average grain size was about 1×10^{-3} cm. which is about the same as the calculated value of b.

* Taylor and Langmuir, reference 25, page 431.

Another striking confirmation of the hypothesis that various crystals of a filament have different work functions and thus emit electrons with greatly varying intensities, is given by pictures of such filaments obtained by means of electron optics.* Figure 21 shows an electron

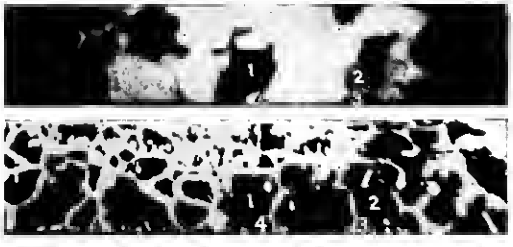


Fig. 21—Electron-micrograph (above) and photo-micrograph (below) of platinum ribbon.

and photo-micrograph † of a portion of a platinum filament. Corresponding crystals have been labeled by 1, 2, 3 and 4. The reader can find more cases of correspondence. Of course, a perfect correspondence is not to be expected since two neighboring crystals may have the same reflection properties for light while the electron emissivities differ and *vice versa*.

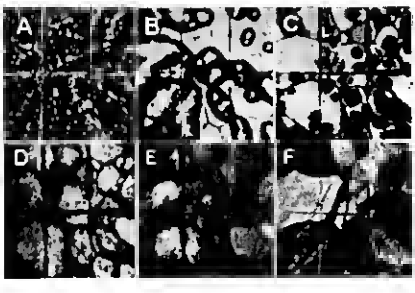


Fig. 22‡—Electron-optical pictures at various stages of heat treatment for a nickel surface coated with oxide.

Figure 22 shows a series of electron-optical pictures taken by W. Knecht [26] of part of a nickel surface covered with BaO and SrO. The

* For an interesting and instructive account of the technique and applications, see the book: *Geometrische Elektronenoptik* by E. Brüche and O. Scherzer, published by J. Springer (Berlin, 1934).

† I am gratefully indebted to Dr. C. J. Davisson and Mr. C. J. Calbick of these laboratories for this figure.

‡ We take this opportunity to thank the publishers of the Annalen der Physik for permission to reproduce this figure from Knecht's article in Annalen der Physik **20**, 180 (1934).

original magnification is 27-fold. Several lines were scratched on the surface. The filament was mounted in a vacuum tube and activated. Picture "a" shows the electron emission from the various granules of the oxide. At this stage the cathode had the characteristic appearance of ordinary oxide coated filaments. The filament was then flashed at a comparatively high temperature for successive intervals of time. After each interval another electron picture was taken. As a result of this treatment the oxide evaporated so that the filament had a metallic appearance. However, there is good reason for believing that metallic barium had been alloyed with the nickel, and the emission was much greater than that from clean nickel. The pictures show that when the oxide has disappeared, different areas emit electrons with greatly different intensities. The shapes and sizes of these areas are strikingly similar to those obtained in ordinary optical pictures. In fact Knecht states that the pattern is that of the nickel crystals.

Numerous other pictures similar to Fig. 22 can be found in the book by Brüche and Scherzer referred to above. Some of these indicate that there is a fine structure non-uniformity inside of a single crystal as well as the non-uniformity between crystals.

The emission from thoriated tungsten has also been investigated by electron optics and while the pictures are not as striking as those for barium on nickel, they prove rather conclusively that the emission varies from one crystal to the next or from one region of a crystal to the next. Finally the emission from surfaces to which no impurity has purposely been added have been investigated. These too show patchy emissions.

It has generally been felt and frequently stated that the emission from clean surfaces, particularly clean tungsten, varied with applied field according to the Schottky equation. While this is true at moderate and high fields, I have never seen data which showed agreement at low fields, but have seen data which showed disagreement. The deviation is less pronounced than it is for thoriated tungsten but in my opinion it is none the less real. This I have interpreted to mean that even in a polycrystalline surface of clean tungsten, individual crystals have work functions which vary by one-tenth or a few tenths of a volt. This interpretation receives strong support from the work of Farnsworth and Rose.[27] They showed that a (111) surface of a single crystal of copper had a contact potential with respect to a (100) surface equal to 0.463 volt. The direction was such that the work function of the (111) surface was 0.463 volt less than that of the (100) surface. Even after heating the crystals to about 900° C. for as much as 1000 hours, the contact potential difference was still 0.378 volt and had

remained practically constant during the last 700 hours. This decrease is ascribed to the formation of new crystal facets, which produce more nearly equal work functions for the two surfaces. That different crystal planes have different work functions also follows from the photoelectric work of Nitzsche [28] on single crystals of zinc. The work function for a surface normal to the hexagonal axis was found to be 3.28 volts while that for a surface parallel to this axis was 3.09 volts. These observed differences in work function of different surfaces of single crystals are quite large enough to account for the deviations from the Schottky law for clean surfaces.

Still another prediction of the patch theory is verified by experiment. In connection with Fig. 15 it was pointed out that at low applied fields something more than half the area is effective in emitting electrons; as the field is increased the effective area decreases until the log i *vs.* $\sqrt{F}$ curves approach the Schottky line when the effective area attains a constant value whose order of magnitude is 0.1. From this it follows that if Richardson lines are obtained for a series of applied potentials, the intercepts or values of log A should decrease as V increases, but should approach a constant value for sufficiently large values of V. Experimental values of log A *vs.* V are given in Fig. 14 of Brattain and Becker's [29] article on thorium on tungsten. They show the predicted trend. Furthermore, the change in log A with V should be most pronounced for large values of μ. It was shown above that the largest values of μ occur in the neighborhood of $f = 0.6$ while near $f = 1.0$, μ is comparatively small. The experimental curves show the largest dependence of log A on V for $f = 0.6$ and only a small dependence for $f = 1$. In this respect too, experiment confirms the theory.

Non-uniformities on the cathode affect the shape of the retarding potential curves as was explained in a previous section. Here too, there is at least qualitative agreement between theory and experiment.

In connection with the analysis of log i *vs.* $\sqrt{F}$ curves we have, in the course of the last five or six years, developed a number of simple methods for computing approximate values of b and μ. We have also proved a number of useful theorems. If and when the interest in this subject warrants it, we intend to publish these methods and theorems.

Checkerboard with Uniform Charge Distribution

While the agreement between experimental log i *vs.* $\sqrt{F}$ curves with theoretical curves based on a hill and valley charge distribution over a checkerboard array is quite good, it is probable that even better

agreement could be obtained if the charge density over the black squares was assumed to be uniform and equal to $\rho + \bar{\mu}$ while over the white squares it was assumed to be uniform and equal to $\rho - \bar{\mu}$. In terms of the previous notation, $\beta = +1$ for all points of the black squares while $\beta = -1$ for all points of the white squares.

If we use the coordinates indicated in Fig. 11A, such a charge distribution can be represented by the following double Fourier series

$$\sigma = \rho + \frac{16\bar{\mu}}{\pi^2} \sum_N \frac{(-1)^{(N-1)/2}}{N} \cos \frac{\pi N x}{b} \sum_K \frac{(-1)^{(K-1)/2}}{K} \cos \frac{\pi K y}{b}, \quad (74)$$

in which N takes on all values, 1, 3, 5, 7, etc., and for each N, K takes on all the values 1, 3, 5, 7, etc. If such a charge distribution is located at a distance l above the surface while its image is located at a distance l below the surface, then the potential energy of an electron due to this double layer is given by

$$P_\sigma/e = -300 \times 4\pi\rho l - \frac{300 \times 65\bar{\mu} l}{\pi} \sum_N \sum_K \frac{(-1)^{(N+K)/2}}{NK} \times \exp\left(\frac{-\pi(N^2+K^2)^{\frac{1}{2}}}{b} z\right) \cos \frac{N\pi x}{b} \cos \frac{K\pi y}{b}. \quad (75)$$

This formula is accurate provided $l/b \ll 1$, which is always fulfilled in any case in which one is likely to be interested. The electric field normal to the surface due to the double layer is

$$\frac{1}{e}\frac{dP_\sigma}{dz} = \frac{300 \times 64\bar{\mu} l}{b} \sum_N \sum_K \frac{(-1)^{(N+K)/2}(N^2+K^2)^{\frac{1}{2}}}{NK} \times \exp\left(\frac{-\pi(N^2+K^2)^{\frac{1}{2}}}{b} z\right) \cos \frac{\pi N}{b} x \cos \frac{\pi K}{b} y. \quad (76)$$

Equations (74), (75) and (76) reduce to the corresponding equations for the hill and valley distribution if $16\bar{\mu}/\pi^2$ is replaced by μ and if only the first term of the double series is used, i.e., if $N = 1$ and $K = 1$. See equations (70) and (71).

Recently Mr. Albert Rose working with Professor L. P. Smith at Cornell University has made calculations for a checkerboard with uniform charge distribution and has compared his computations with experiment. The agreement is as good as we have found and his computed values of b and μ are about the same as ours.

Linford [6] in an excellent review on the external photoelectric effect has shown that a checkerboard distribution of charge or potential satisfactorily accounts for a number of photoelectric phenomena observed with composite surfaces. His equation (42) is almost

identical with equation (75) above if his $V_0 = 8\pi\bar{\mu}l$, his $2j + 1 = N$ and his $2k + 1 = K$. His equation is not quite as general as ours since he deals only with the case for which $\rho = \bar{\mu}$.

Compton and Langmuir * in 1930 presented an interesting discussion of the poor saturation in composite surfaces. They, too, proposed a checkerboard or patch distribution like the one we are discussing, and on page 151 of their paper they give an equation for the potential above such a surface. Unfortunately there is an error in this equation which was pointed out by Linford.[6] They use only a single summation whereas a checkerboard distribution requires a double summation. However, since they use only the first term of their summation and since this first term is the same as the first term of the correct equation, this error is not serious in their case. Their formula, too, is less general than equation (75) since they deal only with the case $\rho = \bar{\mu}$; this is equivalent to assuming that the white squares are clean tungsten and only the black squares are covered.

They reject their patch theory because (1) "to obtain departures from the Schottky curve comparable to those observed, the patches must be assumed to contain many thousands of atoms," and (2) "the patch theory predicts a departure from the Schottky curve which is small with small fields and increases with large fields, whereas exactly the reverse is the actual case." † From what has been said above it is clear that their second objection is really tied up with their first one, for if larger patch sizes are assumed their statement is incorrect and quite good agreement is found with experiment. They assumed a value of $b = 10^{-6}$ cm. whereas the experimental curves require $b \sim 10^{-4}$ cm. They feel that such "extremely non-uniform distributions" or "such large clusters of adsorbed atoms" are "very improbable." One reason for this belief is that "Becker has shown, for example, that a thorium layer at emission temperatures behaves like a two-dimensional gas on the surface."

In my opinion these objections to the checkerboard or patch theory are not well founded. It seems quite natural to me that various crystals on the surface or various crystal facets may have somewhat different adsorptive properties and that consequently different crystals would be covered to different extents with thorium and would thus have different work functions. It is probable that the size of the squares should be comparable to the crystal size which is of the order of 10^{-4} cm. This is still true if thorium migrates over the surface of the tungsten. The successes of the patch theory presented above far outweigh these objections.

* Reference 1, especially pp. 146–160.
† Reference 1, p. 157.

More than this, some of the very data presented by Compton and Langmuir support the generalized checkerboard theory as we have presented it, i.e., taking into account that both the black and the white squares may be covered with thorium but to different extents. On page 155, Compton and Langmuir discuss two log i *vs.* $\sqrt{F}$ curves obtained by Reynolds[30] for thoriated tungsten. Curve A in their Fig. 4 is a "normal" curve while curve C is taken after the surface has been bombarded by positive ions. They state "this bombardment must have roughened the surface and there is evidence that it also fractured the surface layer of tungsten crystals." In our notation this means that b has been decreased because of the roughening and μ has been increased because the amount of thorium removed in some spots was larger than that removed in others. Now the decrease in b should shift the region at which the curve approaches the Schottky line to higher values of F or $\sqrt{F}$; while the increase in μ should result in a steeper curve and should decrease log i_{ao}. (See our Fig. 16.) But this is precisely what curve C does.

Reynolds[30] in discussing this same data says: "The effect of bombardment was a semi-permanent one. Subsequent activation and deactivation by temperature (below 2700°) shifted the curve along the current axis but did not otherwise alter its unique character. Flashing at 2700° K. or higher, where rapid sintering of tungsten is known to take place, destroyed the effect of bombardment and subsequent activation produced normal log i *vs.* $V^{\frac{1}{2}}$ curves." Every detail of this behavior is just what is to be expected on our view; the "semi-permanent" effect is caused by the decrease in b which does not become normal until the damage to the crystals has been repaired by high temperature treatment; the shifting of the curves along the current axis is caused by changes in ρ and μ brought about by activation and deactivation.

On page 156 Compton and Langmuir[1] discuss Kingdon and Langmuir's data for thoriated tungsten at various degrees of activation (θ or f) and at various temperatures. Some of the results are shown in their Fig. 5 which is a plot of T log i *vs.* $\sqrt{F}$. They point out three and only three distinctive features of these curves. All three support the checkerboard theory. They say: "At the highest field-strengths (about 10,000 volts/cm.) the curves are seen to approach the theoretical slope." Our analysis shows that this means that all surfaces have about the same b irrespective of θ and T. This is predicted by our theory since b is determined by the crystal size which is independent of f and T. In discussing curves for a constant f they say: "In every case the departures from the Schottky line become greater as the

temperature is lowered, —." The theory predicts this as shown by Fig. 18. There may, however, be another reason: As T increases μ may decrease. If there are differences in concentration between neighboring crystals, and if the temperature is high enough for migration to occur, Boltzmann's law would require that the difference in concentration should decrease as T increases.

About the third feature they say: "These results indicate that with nearly complete thoriation of the surface ($f = 0.91$) and with a bare surface ($f = 0.00$) the approach to the Schottky curve is fairly close, but relatively large departures occur with incomplete thoriation." This fact which is abundantly confirmed by my experience not only with thorium on tungsten but also with cesium on tungsten, cesium on oxygen on tungsten, and barium on tungsten means that as f increases, μ increases at first, rises to a maximum and then decreases. Such a variation of μ with f is to be expected from the shape of the log i *vs.* f or φ *vs.* f curve which will be discussed more fully later on. As f increases, φ decreases rapidly at first, then more and more slowly until it passes through a minimum when $f = 1$; beyond this point φ increases again. It is natural to expect that Δf, the difference between f for the black and the white squares, should increase with f; Δf is probably nearly proportional to f. From this and the shape of the $\varphi - f$ curve it follows that $\Delta\varphi$, the difference in φ between black and white squares, is small when f is small; as f increases $\Delta\varphi$ increases at first but later on it decreases; when f approaches 1.0, $\Delta\varphi$ approaches 0 and the surface has a uniform work function. Since μ and $\Delta\varphi$ are proportional, μ should vary in the same way. Hence this feature of Compton and Langmuir's curves as well as the first two is entirely in agreement with the predictions based on the checkerboard theory.

Whether the uniform charge distribution or the hill and valley distribution gives better agreement with experiment has not been decided. This, however, is not very important or very pressing. In an experimental filament the distribution is probably neither one nor the other but something in between. Furthermore, it should be emphasized that in an experimental cathode the patches are not all of the same size nor is the contact potential between two neighboring patches a constant; both of these quantities fluctuate about a mean value. Nevertheless I believe that a sufficiently good case has been made out to show that non-uniformities play an important role in many thermionic experiments, and that the checkerboard theory can be used as a powerful tool in the study of adsorption phenomena, where non-uniformities almost always occur.

This analysis of the effect of non-uniformities has brought out that

the work function is not a characteristic of a given substance but rather of a given surface of a given substance. Strictly speaking, one should not talk about the work function of tungsten but rather of the work function of a particular surface of tungsten. This is true even if the surface is clean tungsten.

The Values of the Work Function for Clean Surfaces

The experimental determination of the thermionic work function or the heat function for clean metal surfaces has been the subject of many investigations. In the case of a number of elements, the determinations by different investigators are not in accord. This is due, in most cases, to adsorbed layers of foreign material caused by either poor vacuum conditions or impurities in the metal which have not been eliminated by a proper heat treatment. Although these measurements have been summarized and discussed in other reviews, it seems advisable that the summary be brought up to date and the most probable values selected from the existing data. Since the photoelectric work function is equal to the thermionic work function,[8] the determination by photoelectric methods should also be included.

A summary of the data is shown in Table IV. The values of the photoelectric work function and the thermionic heat function are expressed in volts. The reference for each value is indicated by the superscript. As discussed in an earlier section, the heat function is the slope of a Richardson line. The photoelectric work functions are mostly calculated from the long wave-length limit except in the case of recent determinations which are made by an analysis of the data by Fowler's [31] method. The photoelectric values listed in the table were selected as representative of values for the best outgassing of each element. For a listing of all determinations see Hughes and DuBridge's book. In most cases, the heat function and the thermionic work function differ only by small amounts so that for practical purposes we can compare the photoelectric work function with the heat function. The most probable values of the heat functions tabulated have been chosen from the several determinations.

Recently several attempts have been made to find an empirical relation between the work function and the atomic properties of the elements. Such a correlation, if applicable to all of the metallic elements, would be of value in predicting values of the work function for the cases in which the existing data are inadequate or no data are available. The work of Rother and Bomke [32] gives the best correlation thus far obtained. In their article they have given a summary of the early attempts at a correlation and therefore we will not consider them here.

TABLE IV

Compilation of Values of Photoelectric and Thermionic Work Functions in Volts and the Value of the Heat Function

Element	Photoelectric work function	Thermionic heat function	Probable value of heat function
Ag	(4.58 to 4.75)[61] (4.71 to 4.75)[22]	(4.08)[25]	4.7
Al	(2.99)[39]		3.0
Au	(4.75 to 4.84)[42] (4.86 to 4.92)[22]	(4.32)[25]	4.8
Bi	(4.05)[61] (3.74)[7] (4.37)[42]		4.1
C	(4.72)[28] (4.81)[46]	(3.93)[26]	4.7
Ca	(2.76)[60] (3.20)[45]	(3.02)[29] (2.24)[19]	3.2
Cb		(3.96)[57]	4.0
Cd	(4.07)[2]		4.1
Ce	(2.84)[45]		2.8
Co	(4.12 and 4.25)[4]		4.2
Cr	(4.60)[35]		4.6
Cs	(1.67)[50]	(1.81)[34]	1.8
Cu	(4.49)[61] (4.08)[33]	(3.85)[55] (4.33)[25]*	4.1
Fe	(4.72)[3] (4.77)[24]	(4.04)[33] (4.77)[52] (4.04)[11]	4.7
Ge	(4.85)[60]		4.9
Hf		(3.53)[64]	3.5
Hg	(4.53)[16, 27, 47]		4.5
K	(1.77)[31] (2.0)[32]		1.8
Li	(2.21)[30]		2.2
Mg	(approx. 2.43)[5]		2.4
Mo	(4.15)[18]	(4.08)[37] (4.39)[58] (4.59)[38] (4.44)[26] (4.38)[63] (3.48)[40] (4.14–4.17)[16] (4.32)[1]	4.3
Na	(1.80)[41] (2.25)[44, 53] (1.94)[59]		1.9
Ni	(5.01)[24]	(4.41)[33] (2.77)[51] (4.31)[57] (4.63)[11] (5.03)[25]	5.0
Os		(4.7)[6]	
Pb	(3.50)[44] (3.97)[36] (4.14)[28] (3.97)[39]		4.0
Pd	(4.97)[17]	(4.99)[17]	4.98
Pt	(6.30)[14]	(6.27)[15] (5.40)[56] (5.93)[49]*	6.0
Rb	(1.82)[50]		1.8
Re	(approx. 5.0)[21]		5.0
Rh	(4.95 to 4.57)[12]	(4.58)[12]	4.6
Sb	(4.02)[20]		4.0
Se	(4.62)[23]		4.6
Sn	(β4.50; γ4.38; liq. 4.24)[25] (β4.39)[62]		4.4
Sr	(2.06)[13]		2.1
Ta	(4.10–4.14)[29] (4.12–4.19)[5] (4.12)[45]	(4.2)[38] (4.51)[33] (4.18)[54] (4.07)[26] (4.04)[40]	4.10
Th	(3.34)[44] (3.57)[28] (3.38)[45]	(3.35)[34]	3.4
U	(3.63)[45]		3.6
W	(4.69 and 4.54)[59] (4.60)[45]	(4.52)[9] (4.53)[11]*	4.52
Zn	(3.61)[26] (3.68)[49] (3.08)[28] (3.32 and 3.57)[10]		3.3
Zr	(3.73)[48]	(4.13)[64]	4.1

* For complete listing see S. Dushman, *Rev. Mod. Phys.* **2**, 381 (1930).

[1] A. J. Ahearn, Phys. Rev. **44**, 277 (1933).
[2] H. Bomke, Ann. d. Physik **10**, 579 (1931).
[3] A. B. Cardwell, Proc. Nat. Acad. Sci. **14**, 439 (1928).
[4] A. B. Cardwell, Phys. Rev. **38**, 2033 (1931).
[5] A. B. Cardwell, Phys. Rev. **38**, 2041 (1931).
[6] R. J. Castman and W. S. Huxford, Phys. Rev. **43**, 811 (1933).
[7] Chien Cha, Phil. Mag. **49**, 262 (1925).
[8] Cooke and Richardson, Phil. Mag. **25**, 624 (1913); **26**, 472 (1913).
[9] C. J. Davisson and L. H. Germer, Phys. Rev. **20**, 300 (1922); **24**, 666 (1924).
[10] J. H. Dillon, Phys. Rev. **38**, 408 (1931).
[11] W. Distler and G. Monch, Zeits. f. Physik **84**, 271 (1933).
[12] E. H. Dixon, Phys. Rev. **37**, 60 (1931).
[13] R. Dopel, Zeits. f. Physik **33**, 237 (1925).
[14] L. A. DuBridge, Proc. Nat. Acad. Sci. **12**, 162 (1926).
[15] L. A. DuBridge, Phys. Rev. **32**, 961 (1928).
[16] L. A. DuBridge and W. W. Roehr, Phys. Rev. **39**, 99 (1932).
[17] L. A. DuBridge and W. W. Roehr, Phys. Rev. **42**, 52 (1932).
[18] H. K. Dunn, Phys. Rev. **29**, 693 (1927).
[19] S. Dushman, Phys. Rev. **21**, 623 (1923).
[20] Dushman, Rowe, Ewald and Kidner, Phys. Rev. **25**, 338 (1925).
[21] A. Engelmann, Ann. d. Physik **17**, 185 (1933).
[22] R. H. Fowler, Phys. Rev. **38**, 45 (1931).
[23] G. W. Fox and R. M. Bowie, Phys. Rev. **44**, 345 (1933).
[24] G. N. Glasoe, Phys. Rev. **38**, 1490 (1931).
[25] A. Goetz, Physik. Zeits. **24**, 377 (1923); **26**, 206 (1925); Zeits. f. Physik **42**, 329 (1927); **43**, 531 (1927).
[26] A. Goetz, Phys. Rev. **33**, 373 (1929).
[27] W. B. Hales, Phys. Rev. **32**, 950 (1928).
[28] R. Hamer, J. Opt. Soc. Am. **9**, 251 (1924).
[29] F. Horton, Phil. Trans. **A207**, 149 (1907).
[30] A. L. Hughes, Phil. Trans. **A212**, 205 (1912).
[31] H. E. Ives and A. L. Johnsrud, Astrophys. J. **60**, 231 1924).
[32] H. E. Ives, Jour. Opt. Soc. Am. **8**, 551 (1924).
[33] Jentzsch, Ann. d. Physik **27**, 148 (1908).
[34] K. H. Kingdon, Phys. Rev. **25**, 892 (1925).
[35] Kösters, Zeits. f. Physik **66**, 825 (1930).
[36] I. Langmuir, Phys. Rev. **2**, 450 (1913); Physik. Zeits. **15**, 516 (1914).
[37] I. Langmuir, Trans. Am. Electrochem. Soc. **29**, 125 (1916).
[38] H. Lester, Phil. Mag. **31**, 197 (1916).
[39] Lukirsky and Prilezaev, Zeits. f. Physik **49**, 236 (1928).
[40] M. J. Martin, Phys. Rev. **33**, 991 (1929).
[41] R. A. Millikan, Phys. Rev. **7**, 355 (1916).
[42] L. W. Morris, Phys. Rev. **37**, 1263 (1931).
[43] T. J. Parmley, Phys. Rev. **30**, 656 (1927).
[44] Pohl and Pringsheim, Verh. d. Phys. Grs. 1911–14.
[45] Rentschler, Henry and Smith, Rev. Sci. Inst. **3**, 794 (1932).
[46] N. B. Reynolds, Phys. Rev. **35**, 158 (1930).
[47] Roller, Phys. Rev. **36**, 738 (1930).
[48] S. C. Roy, Proc. Roy. Soc. **A112**, 599 (1926).
[49] R. W. Sears, Unpublished work.
[50] E. F. Seiler, Astrophys. Jour. **52**, 129 (1920).
[51] W. Schlichter, Ann. d. Physik **47**, 573 (1915).
[52] G. Siljeholm, Ann. d. Physik **10**, 178 (1931).
[53] W. Souder, Phys. Rev. **8**, 310 (1916).
[54] H. J. Spanner, Ann. d. Physik **75**, 609 (1924).
[55] E. R. Stoekle, Phys. Rev. **8**, 534 (1916).
[56] H. L. Van Velzer, Phys. Rev. **44**, 831 (1933).
[57] H. B. Wahlin and L. O. Sordahl, Phys. Rev. **45**, 886 (1934).
[58] Wehnelt and Seiliger, Zeits. f. Physik **38**, 443 (1926).
[59] A. H. Warner, Phys. Rev. **38**, 1871 (1931).
[60] G. B. Welch, Phys. Rev. **31**, 709 (1928).
[61] S. Werner, Upsala Univ., 1914.
[62] R. P. Winch, Phys. Rev. **37**, 1269 (1931).
[63] C. Zwikker, Proc. Amst. Acad. Sci. **29**, 792 (1926).
[64] C. Zwikker, Physik. Zeits. **30**, 578 (1929).

In the preceding sections the thermionic work function W was shown to be equal to $P - K$ where P is the difference in potential energy between an electron at rest inside and outside of the metal and K is given by equation (16). If we assume that there is one free electron per atom in the metal for all elements, then

$$K/e = 25.9(D/M)^{\frac{1}{3}}, \tag{77}$$

where D is the density of the metal and M the atomic weight.

From values of K/e given by equation (77) and experimental values of W/e, Rother and Bomke calculated P/e for a number of elements. Their values of P/e were said to be in accord with the empirical equations

$$P/e = 12.6(Dz/M)^{\frac{1}{3}} \text{ for some elements} \tag{78}$$

and

$$P/e = 16.3(Dz/M)^{\frac{1}{3}} \text{ for all other elements,} \tag{79}$$

where z is the maximum chemical valence of the element.

We have computed values of K/e from equation (77) and with the most probable values of W/e from Table IV have determined the probable values of P/e. Since the work function and the heat function differ by only small amounts, it is justifiable to use the heat functions for W/e. To test equations (78) and (79) we have plotted log P/e *vs.*, log (Dz/M) in Fig. 23. According to equations (78) and (79), the points should fall on two straight lines in this plot. The two lines are shown in the figure and have a slope of $\frac{1}{3}$. The values of z used in this plot are those given by Rother and Bomke. The points lie in the general neighborhood of the lines but there is no clear indication of a division into two groups. The deviations in about half of the cases are larger than the possible experimental error.

Bomke [33] has recently found that his values of P/e (from calculated K/e and experimental W/e) plotted against the compressibility gave a smooth curve. The equation of this curve was

$$P/e = 0.30k^{-\frac{1}{3}}, \tag{80}$$

where k is the compressibility. Unfortunately he plotted his data on a linear scale and most of the points on his plot were clustered near one axis where the curve was steep, making it difficult to estimate the deviations. A plot of log P/e *vs.* k which is similar to Fig. 23 showed that the deviations were of the same order of magnitude as the deviations in Fig. 23 previously discussed. Hence values of P/e calculated

from equation (80) will only be approximate. The approximation is about the same as computing P/e from equation (78) or (79). Equation (80) has the advantage that P/e is given by a single function.

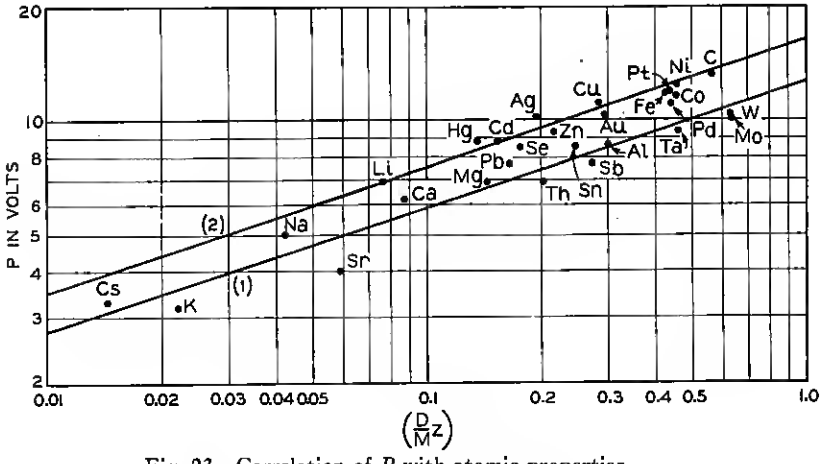


Fig. 23—Correlation of P with atomic properties.

Chittum [34] has related the work function to the bulk modulus of compressibility in a different manner than that used by Bomke. His computed values of the work function deviate from the experimental values by about the same amount as do the values computed from equations (77) and (80).

From the fact that metals with large atomic spacing have a low P while those with small atomic spacing have a high P, it might be expected that P would depend on the spacing of the atoms in the surface layer and that different faces of a single crystal would have different work functions. In fact, Farnsworth and Rose [27] have shown that the contact potential for different faces of a single crystal of Cu varies by about 0.4 volt; and Nitzsche [28] finds the photoelectric work function of two planes of a single crystal of zinc different by about 0.2 volt. Now the values of (D/M) or of the cubic compressibility used in the above calculations do not take into consideration any dependence on the crystal face exposed and, therefore, we would not expect P to be a single-valued function of these properties. This fact may explain the failure to correlate exactly the body properties of the metal with P or the work function. It is quite likely that a better correlation exists between the work function or P and the atomic spacing which prevails on the crystal faces which develop when a metal is heated in a vacuum.

F. Currents Limited by Space Charge

Thus far we have only considered the effect of the surface and applied fields on the number of electrons that escape from a cathode and reach the anode at a particular temperature. However, an electron traveling from the cathode to the anode is also subjected to a field due to all of the electrons in the space between the electrodes. If the electron density in this space is large enough, the current that reaches the anode will be determined by these charges rather than by the work function and temperature of the cathode. The current is then said to be limited by space charge. If, on the other hand, the applied potential is raised to a sufficiently high value, the current is no longer limited by the charges in the space but is then determined by the work function and temperature of the cathode. The current is then said to be saturated or limited by emission. The space charge and saturated emission regions are illustrated by curve 2 in Fig. 24 which is a plot of log i *vs.* log V. In the region to the left of point A, the current is limited by space charge and increases rapidly with the applied potential. To the right of A, the current is limited by emission. Curve 2 has been calculated from equations that will be discussed later. A sharp break point is indicated at A, whereas experimental curves usually show a gradual transition. This gradual transition is due to non-uniformities in work function.

When the current is limited by space charge, the charges in the space increase the height of the potential barrier which electrons must cross in traveling from cathode to anode. The current is determined primarily by the applied potential and electrode geometry and secondarily by the temperature of the cathode and magnitude of the saturated emission. The problem of relating the current to these quantities is very difficult but has been solved on the basis of certain simplifying assumptions for several forms of electrodes by Child,[38] Schottky,[39] Epstein,[40] Fry,[41] Langmuir [42–45] and others. These assumptions together with the solutions will be summarized in this section.

In all of these solutions it is assumed that the maximum of the potential hill which is due to the surface forces and the applied potential, occurs right at the cathode surface. Actually, in the absence of space charge, the maximum in the work distance curve occurs at a small but finite distance from the surface, about 3×10^{-6} cm. for the image equation with moderate applied fields. The space within this distance has a much larger density of electrons than if the potential had its maximum value at the surface. The above assumption, therefore, neglects the influence of these excess charges on the space charge.

This has been justified by Schottky [36] and Laue [37] who concluded that the effect of these charges is negligible. For convenience, the zero of potential is taken not inside the metal but at a point where the electrons

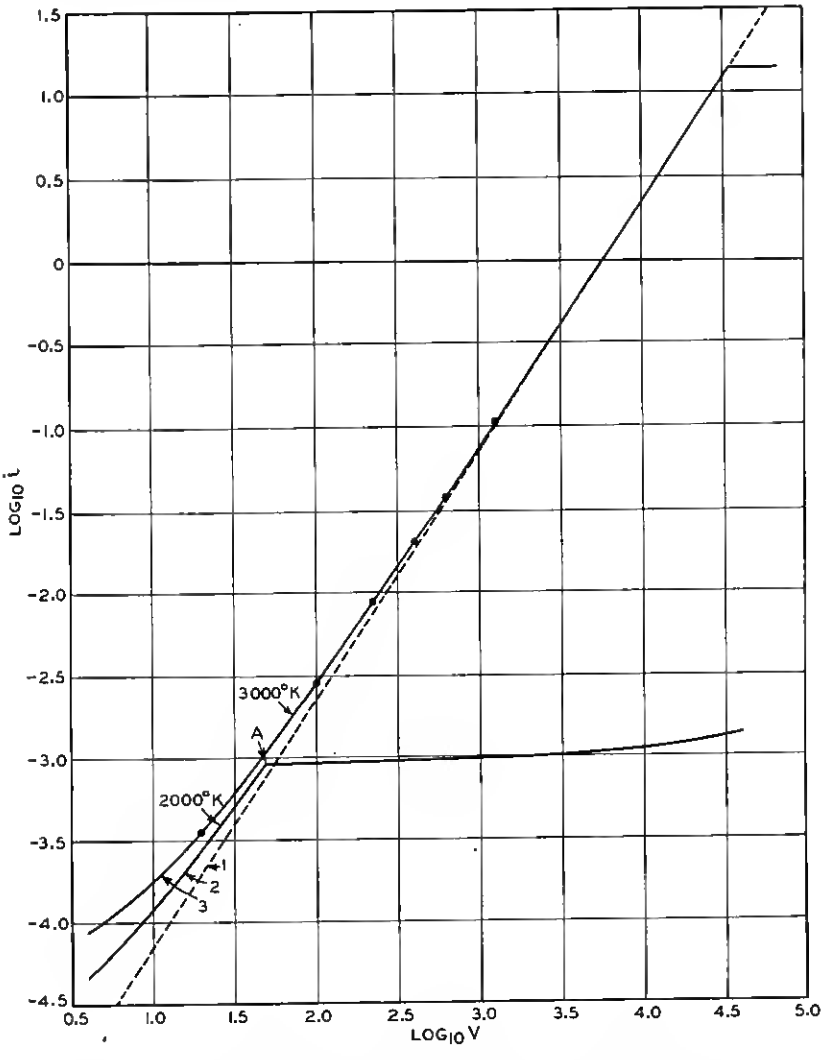


Fig. 24—Currents between parallel plates limited by space charge.

have just overcome the surface forces. Because of the assumption made above, this point is taken at the cathode surface.

In order to obtain a solution, some assumption regarding the velocity

distribution of the emitted electrons must be made. The simplest assumption is that the electrons are emitted from the cathode with zero velocity. From this assumption and the assumption discussed in the preceding paragraph it follows that the potential maximum always occurs at the cathode surface and the emitted electrons are accelerated everywhere in their path between cathode and anode. The current to the anode is determined by the potential distribution in the space.

A better assumption is that the electrons are emitted with a Maxwellian velocity distribution. For this case the potential maximum occurs at some distance from the cathode. The position and value of the maximum depends upon the work function and temperature of the cathode, the geometry of the electrodes and the applied potential. In order that an electron shall reach the anode, its initial velocity normal to the surface must correspond to an energy which is equal to or greater than the potential maximum.

For parallel plates and cylindrical electrodes the following solutions have been obtained for the two assumptions in regard to the velocity distribution of the emitted electrons.

Electrons Emitted with Zero Velocity

For infinite parallel plates:

$$i = (\sqrt{2}/9\pi)(e/m)^{\frac{1}{2}}(V^{\frac{3}{2}}/x^2) = 2.33 \times 10^{-6}(V^{\frac{3}{2}}/x^2), \qquad (81)$$

where i is the current to the anode in amp. per cm.^{2}; x the distance in cm. between cathode and anode; and V is the applied potential in volts corrected for the contact potential; e and m are the charge and mass of the electron, respectively.

For long coaxial cylinders:

$$i = (2\sqrt{2}/9)(e/m)^{\frac{1}{2}}(V^{\frac{3}{2}}/R\beta^2) = 1.48 \times 10^{-5}(V^{\frac{3}{2}}/R\beta^2), \qquad (82)$$

β is a function of R/r_0 where R and r_0 are the radii of anode and cathode, respectively. Table V shows a few values of β^2 as a function of R/r_0

TABLE V

VALUES OF β^2

R/r_0	1.0	2.0	5.0	7.0	10.0	20.0	40.0
β^2	0.000	0.279	0.767	0.887	0.978	1.072	1.095
R/r_0	70	100	200	400	1000	5000	∞
β^2	1.088	1.078	1.056	1.036	1.017	1.002	1.000

taken from a table given by Langmuir and Blodgett.[44]

The equation for coaxial cylinders only applies to an equipotential cathode. Ordinarily the cathode is a filament and the potential varies along its length because of the heating current. It is of interest to examine the modification of the equation due to this effect. It can be shown that for $V_p > V_f$

$$i = (2\sqrt{2}/9)(e/m)^{\frac{1}{2}}(V_p^{\frac{3}{2}}/R\beta^2) \times [1 - 3V_f/4V_p + 3/24(V_f/V_p)^2 \cdots], \quad (83)$$

where V_p is the applied potential between anode and negative end of the filament and V_f is the total potential drop along the filament.

For the case in which $V_p < V_f$

$$i = (2\sqrt{2}/9)(e/m)^{\frac{1}{2}}(2V_p^{5/2}/5R\beta^2 V_f) = 5.92 \times 10^{-6} V_p^{5/2}/RV_f. \quad (84)$$

For concentric spheres:

$$i = (4\sqrt{2}/9)(e/m)^{\frac{1}{2}}(V^{\frac{3}{2}}/\alpha^2) = 2.96 \times 10^{-5} V^{\frac{3}{2}}/\alpha^2, \quad (85)$$

where α^2 is a function of R/r_0 and has been tabulated by Langmuir and Blodgett.[45] α^2 increases with R/r_0. For $R/r_0 = 5.0$, $\alpha^2 = 1.141$; for $R/r_0 = 10$, $\alpha^2 = 1.777$; for $R/r_0 = 100$, $\alpha^2 = 3.652$.

Electrons Emitted with Maxwellian Velocity Distribution

For infinite parallel plates the space charge limited current is given by

$$i = (\sqrt{2}/9\pi)(e/m)^{\frac{1}{2}}((V - V_m)^{\frac{3}{2}}/(x - x_m)^2) \times [1 + 2.66(kT/(V - V_m)e)^{\frac{1}{2}}]$$
$$= 2.33 \times 10^{-6}((V - V_m)^{\frac{3}{2}}/(x - x_m)^2) \times [1 + 2.48 \times 10^{-2}(T/(V - V_m))^{\frac{1}{2}}]. \quad (86)$$

In this equation

$$V_m = (-2.3Tk/e) \log (i_s/i) = -1.98 \times 10^{-4} T \log (i_s/i) \quad (87)$$

and

$$x_m = 1.092 \times 10^{-6} T^{3/4} \zeta_1 / i^{\frac{1}{2}}, \quad (88)$$

where V is the potential applied between cathode and anode corrected for the contact potential; V_m is the value of the potential maximum measured with respect to the zero of potential previously defined; x the distance between cathode and anode; x_m the distance from the cathode to the potential maximum; T the temperature of the cathode; and i_s the value of the saturation electron emission. ζ_1 is a function of ln (i_s/i). Table VI gives a few values of ζ_1 as a function of ln

TABLE VI

VALUES OF ζ_1.

ln (i_s/i)	0.00	0.30	0.60	1.00	1.60	2.40
ζ_1	0.000	0.979	1.312	1.600	1.881	2.117
ln (i_s/i)	3.4	4.5	7.0	10.0	15.0	25.0
ζ_1	2.293	2.404	2.511	2.544	2.553	2.554

(i_s/i). The values listed in the table were selected from a more extensive table given by Langmuir.[43] From equation (87) it follows that space charge acts as if the work function of the surface were increased by V_m.

An expression for i as a function of V could be obtained by eliminating V_m and x_m between equations (86), (87) and (88). Because of the nature of these equations an analytical expression for i cannot be given. However, for any temperature and electrode spacing it is possible to calculate i as a function of V. The effect of introducing the Maxwellian distribution of velocities can be seen by comparing curves calculated from equations (86), (87) and (88) with equation (81). Such a comparison is made in Fig. 24. Equation (81) gives a straight line with a slope of 3/2 on such a plot and is represented by curve 1. Curves 2 and 3 were calculated from equations (86), (87) and (88) for parallel plates of tungsten spaced 1 cm. apart at 2000 and 3000° K., respectively. The introduction of the Maxwellian velocity distribution causes the currents to be somewhat higher than predicted by the simple 3/2 power law, especially at low applied potentials. Furthermore, at applied potentials considerably less than necessary for saturation, the slope of log i *vs.* log V is less than 3/2. Near the break point, the slope is practically 3/2.

For long coaxial cylinders, Schottky [39] and Langmuir [43, 44] have pointed out that the effect of introducing the Maxwellian velocity distribution is less important than its introduction in the plane parallel case. Langmuir [43] has discussed an approximate formula for this case.

Effect of Fermi-Dirac Velocity Distribution

The introduction of the newer theory that the free electrons in the metal have a Fermi-Dirac velocity distribution requires no modification of the space charge equations deduced on the assumption of a Maxwellian distribution. This is because of the fact that the electrons which escape across the potential barrier at the surface have a Maxwellian distribution, as was shown in connection with Fig. 1. In this connection it is of interest to compare some calculations by Bartlett [46] as-

suming a Fermi-Dirac distribution in the metal at 3000° K. with calculations based on a Maxwellian distribution. This comparison also is shown in Fig. 24, curve 3. The circles were taken from a curve by Bartlett and the line was calculated by equations (86), (87) and (88). The two calculations agree, thus indicating that the assumption of a Fermi distribution in the metal leads to the same result as the assumption of a Maxwellian distribution.

G. Miscellaneous Topics

In order not to lengthen unduly this review we have omitted a discussion of a number of topics. Such topics have either been adequately treated in the reviews and books referred to in the introduction or else no significant advances have been made recently. Some of these topics are: Secondary electron emission, high field emissions, thermionics as related to photoelectricity and contact potential,* and cooling and heating effects accompanying the emission or absorption of electrons. In connection with the last topic we feel that a critical analysis of how the quantities determined by experiment are related to the work function and heat function should be made. Most of these experiments were performed before the day of the Fermi-Dirac-Sommerfeld contributions and should thus be reinterpreted.

Acknowledgments

It gives me pleasure to acknowledge the help I have received in the preparation of this article from my colleagues at the laboratories. I acknowledge in particular the benefit of numerous discussions with Dr. C. J. Davisson and Dr. W. H. Brattain and Mr. R. W. Sears. Professor V. Rojansky, now at Union College, is responsible for some of the more complicated equations used in the checkerboard theory. Mr. R. W. Sears deserves a great deal of credit for his painstaking efforts in preparing the figures and tables and in assembling some of the data.

References

1. K. T. Compton and I. Langmuir, *Rev. Mod. Phys.*, **2**, 123 (1930).
2. S. Dushman, *Rev. Mod. Phys.*, **2**, 381 (1930).
3. A. L. Reimann, *Thermionic Emission*, Aberdeen University Press, Aberdeen, Scotland, or John Wiley & Sons, Inc. (New York, 1934).
4. Müller-Pouillets, *Lehrbuch der Physik*, Vol. IV, F. Vieweg, Braunschweig, 1934.
5. Wien Harms, *Handbuch der Experimentalphysik*, Vol. XIII, Part 2, Akademische Verlagsgesellschaft (Leipzig, 1928).
6. L. B. Linford, *Rev. Mod. Phys.*, **5**, 34 (1933).
7. A. L. Hughes and L. A. DuBridge, McGraw-Hill Book Co. (New York, 1932).
8. J. A. Becker and W. H. Brattain, *Phys. Rev.*, **45** ,694 (1934).
9. P. W. Bridgman, *The Thermodynamics of Electrical Phenomena in Metals*, Macmillan Co. (New York, 1934).
10. S. Dushman, *Phys. Rev.*, **21**, 623 (1923).

* For a critical discussion of this relationship, see Becker and Brattain.[8]

11. A. Sommerfeld, *Zeits. f. Physik*, **47**, 1, 43 (1928); *Naturwiss.*, **15**, 825 (1927).
12. A. Sommerfeld and N. H. Frank, *Rev. Mod. Phys.*, **3**, 1 (1931).
13. R. H. Fowler, *Statistical Mechanics*, University Press (Cambridge, 1929).
14. N. F. Mott, *Proc. Phys. Soc.*, **46**, 680 (1934).
15. W. Hume-Rothery, *The Metallic State*, Oxford University Press (London, 1931).
16. Nordheim, *Physik. Zeits.*, **30**, 177 (1929).
17. L. Tonks, *Phys. Rev.*, **38**, 1030 (1931).
18. Bowden and Rideal, *Proc. Roy. Soc.*, **A120**, 59 (1928); Bowden, *Proc. Roy. Soc.*, **A125**, 446 (1929).
19. W. Schottky, *Ann. d. Physik*, **44**, 1011 (1914).
20. L. H. Germer, *Phys. Rev.*, **25**, 795 (1925).
21. C. J. Davisson, *Phys. Rev.*, **23**, 299 (1924).
22. A. Demski, *Physik. Zeits.*, **30**, 291 (1929).
23. W. B. Nottingham, *Phys. Rev.*, **41**, 793 (1932).
24. J. A. Becker and D. W. Mueller, *Phys. Rev.*, **31**, 431 (1928).
25. J. B. Taylor and I. Langmuir, *Phys. Rev.*, **44**, 423 (1933).
26. W. Knecht, *Ann. d. Physik*, **20**, 161 (1934).
27. H. E. Farnsworth and B. A. Rose, *Proc. Nat. Acad. Sci.*, **19**, 777 (1933); B. A. Rose, *Phys. Rev.*, **44**, 585 (1933).
28. A. Nitzsche, *Ann. d. Physik*, **14**, 463 (1932).
29. W. H. Brattain and J. A. Becker, *Phys. Rev.*, **43**, 428 (1933).
30. N. B. Reynolds, *Phys. Rev.*, **35**, 158 (1930).
31. R. H. Fowler, *Phys. Rev.*, **38**, 45 (1931).
32. F. Rother and H. Bomke, *Zeits. f. Physik*, **86**, 231 (1933).
33. H. Bomke, *Zeits. f. Physik*, **90**, 542 (1934).
34. J. F. Chittum, *Jour. Phys. Chem.*, **38**, 79 (1934).
35. W. Schottky, *Physik. Zeits.*, **15**, 872 (1914).
36. W. Schottky, *Physik. Zeits.*, **15**, 624, 872 (1914).
37. M. v. Laue, *Jahrb. d. Radioakt.*, **15**, 205 (1918); *Berl. Ber.*, **32**, 334 (1923).
38. C. D. Child, *Phys. Rev.*, **32**, 492 (1911).
39. W. Schottky, *Physik. Zeits.*, **15**, 526 (1914).
40. P. S. Epstein, *Verh. d. Phys. Ges.*, **21**, 85 (1919).
41. T. C. Fry, *Phys. Rev.*, **17**, 441 (1921).
42. I. Langmuir, *Phys. Rev.*, **2**, 402 (1913).
43. I. Langmuir, *Phys. Rev.*, **21**, 419 (1923).
44. I. Langmuir and K. B. Blodgett, *Phys. Rev.*, **22**, 347 (1923).
45. I. Langmuir and K. B. Blodgett, *Phys. Rev.*, **24**, 49 (1924).
46. R. S. Bartlett, *Phys. Rev.*, **37**, 959 (1931).